E-ISSN: XXXX - XXXX
Networked Knowledge and Research in Informatic

Networked Knowledge and Research in Informatic
Vol. 1 No. 1 (2026) : 09-18
Available online at : https://ejournal.unuja.ac.id/index.php/NKRI

Clustering Divorce Cases at Kraksaan Religious Court Using K-Means

Zainal Arifin [1]
[1] Universitas Nurul Jadid, Indonesia
zainalarifin@unuja.ac.id [1]

Doi
Submitted: 01 Sep 2026 Revision: 15 Sep 2026 Accepted: 25 Sep 2026 Published: 30 Sep 2026

Abstract Divorce cases are social and legal issues that require systematic data mapping so that relevant institutions can understand variations in case characteristics across regions. This study aims to cluster divorce cases at the Kraksaan Class 1A Religious Court based on subdistrict, age difference between spouses, and factors causing divorce using the K-Means Clustering algorithm. Primary data were obtained through observation, interviews, and documentation of divorce case records from 2022 to June 2024. The dataset consisted of 2,424 case records, which underwent data selection, data cleaning, categorical-to-numerical transformation, and data aggregation by subdistrict. The variables used included the frequency of age-difference categories and the frequency of divorce-causing factors in each subdistrict. The clustering process used three clusters to map high, medium, and low divorce-case profiles. The results show that Cluster 0 represents the group with the highest divorce-case profile and consists of Maron Subdistrict. Cluster 1 represents the low divorce-case group and consists of nine subdistricts, while Cluster 2 represents the moderate divorce-case group and consists of fourteen subdistricts. Internal evaluation using the Davies-Bouldin Index produced a value of 0.6611. This value indicates reasonably good cluster separation because a smaller DBI value indicates a more compact and better-separated cluster structure. The clustering results were integrated into a Streamlit application to facilitate data display, preprocessing, cluster results, and analytical visualization.

Keywords *Clustering; Davies-Bouldin Index; Divorce Cases; K-Means; Streamlit*

1. INTRODUCTION

Divorce is a social and legal issue that can affect individuals' psychological conditions, family continuity, child welfare, and social relationships within communities. Each divorce case has different characteristics, including the couple's area of residence, age difference, and factors contributing to the divorce. Therefore, divorce case data need to be analyzed systematically so that hidden patterns can be identified and used as supporting information for courts and related institutions (Amrina, 2022). The Kraksaan Class 1A Religious Court handles various family-related cases, including divorce initiated by wives and divorce initiated by husbands. The increasing number of case records over time makes it difficult to identify patterns through manual examination alone. In fact, case data have the potential to provide important information regarding subdistricts with certain case concentrations, patterns of age differences between spouses, and the dominant causes of divorce. However, such information cannot be optimally obtained without the support of data analysis methods.

Data mining is an approach used to discover hidden patterns, relationships, and knowledge from large datasets. One of the main techniques in data mining is clustering, which is the process of grouping data based on similarity without using predefined class labels (Mardi, 2017; Samosir et al., 2021). In the context of divorce cases, clustering can be used to map subdistricts with similar case patterns so that differences in characteristics across regions can be identified more clearly. K-Means is a non-hierarchical clustering method that is widely used because it is relatively simple, fast, and easy to implement. The algorithm works by assigning a number of cluster centers or centroids, calculating the distance of each data point to the centroids, updating the cluster centers, and repeating the process until the cluster structure becomes stable. Previous studies have used K-Means to group divorce cases based on geographical areas and case characteristics, including studies in Simalungun, Jambi City, Kuningan Regency, Bekasi City, and several other religious court jurisdictions (Rahayu et al., 2022; Sa'dah et al., 2019; Sopyan et al., 2022; Yanti, 2021).

Although previous studies have shown that K-Means can be used for divorce-case mapping, differences remain in research locations, data volume, variables, and result presentation. This study uses primary divorce-case data from the Kraksaan Class 1A Religious Court, focusing on subdistricts, age differences between spouses, and factors causing divorce. In addition, the clustering results are not only presented as calculation outputs but are also implemented in a Streamlit application so that administrators can review the data and visualized results more conveniently.

Clustering quality needs to be evaluated using an appropriate metric. The Davies-Bouldin Index (DBI) is used to measure the compactness of data within clusters and the separation between clusters. A lower DBI value indicates that data within a cluster are closer to one another and more clearly separated from other clusters (Gie & Jollyta, 2020; Sopyan et al., 2022). In clustering studies, DBI should not be interpreted as classification accuracy; instead, it serves as an internal evaluation measure of cluster quality.

This study aims to apply the K-Means Clustering algorithm to group divorce cases at the Kraksaan Class 1A Religious Court. The additional objectives are to

evaluate cluster quality using the Davies-Bouldin Index and to implement the clustering results in a Streamlit application. The findings are expected to provide supporting information for understanding the distribution and patterns of divorce cases across subdistricts within the jurisdiction of the Kraksaan Religious Court.

2. RESEARCH METHODS

This study employed a quantitative approach using data mining methods. The research stages included data collection, data selection, data cleaning, data transformation, cluster formation using K-Means, evaluation using the Davies-Bouldin Index, and implementation of the results in a Streamlit application.

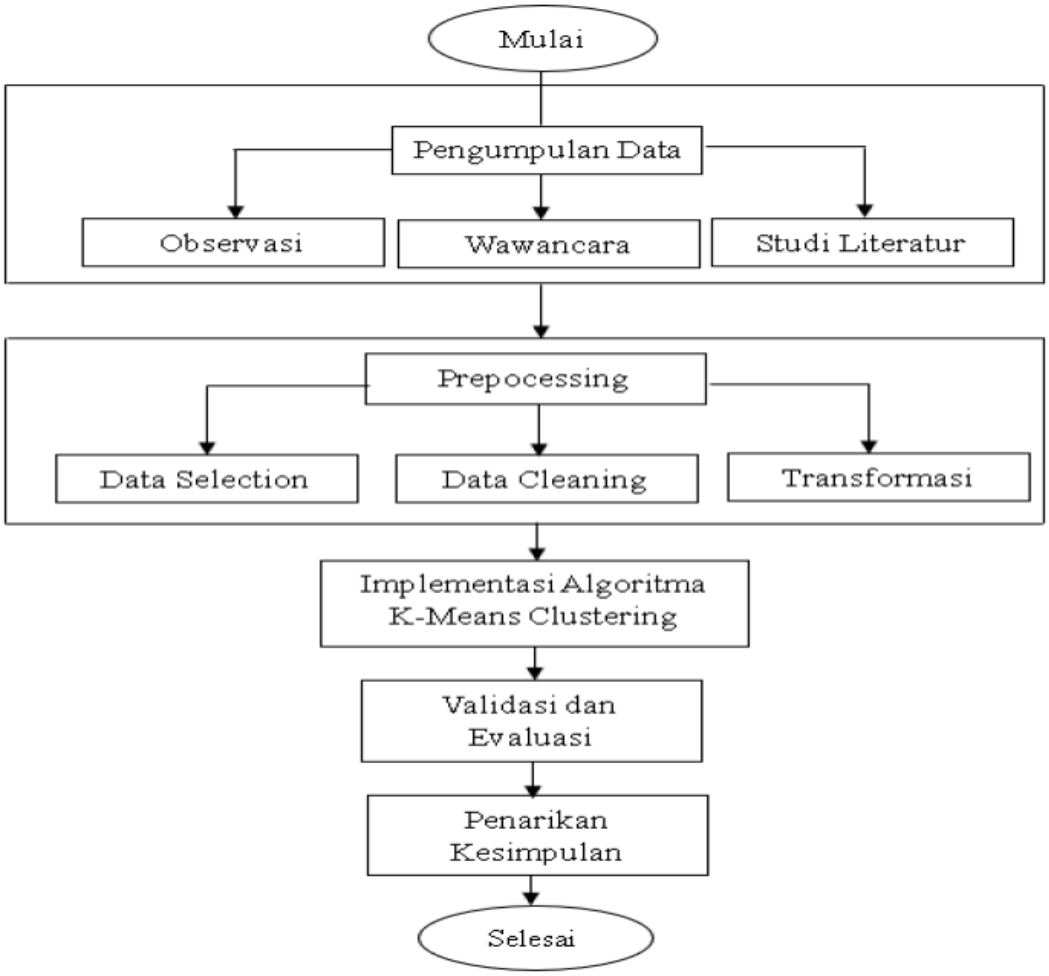


Figure 1. Research Flow Diagram

2.1. Data Collection

The research data consisted of primary data obtained from the Kraksaan Class 1A Religious Court. Data collection was conducted through observation, interviews with relevant staff, and documentation of divorce case records. The data included divorce cases from 2022 to June 2024, with an initial total of 2,424 records. Personal identity information, such as the names of petitioners and respondents, was not used in the clustering process. The analysis only used attributes relevant to divorce-pattern mapping, namely subdistrict, spouses' ages, and factors causing divorce.

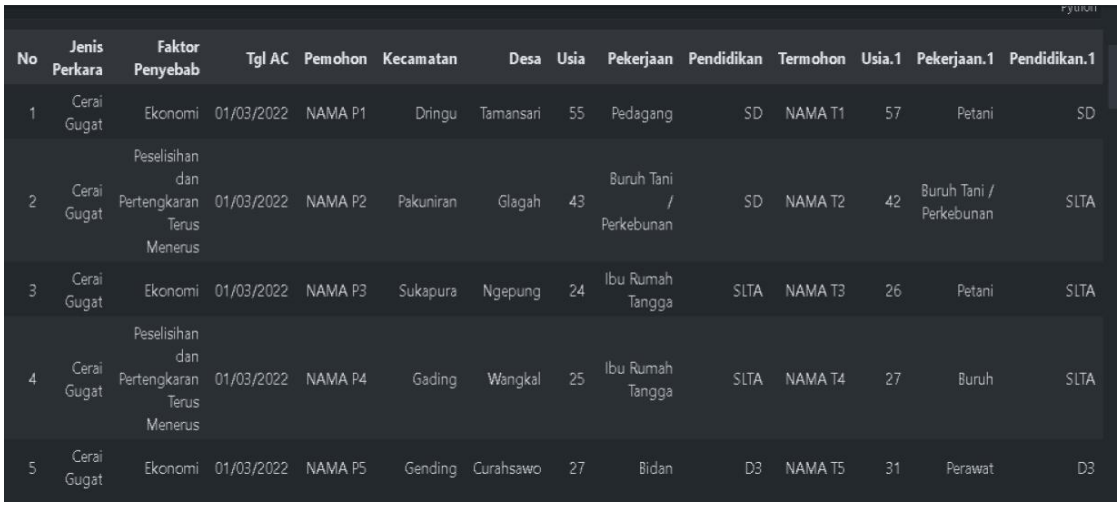
Table 1. Research Dataset Summary

Component	Description

Data source	Kraksaan Class 1A Religious Court
Data type	Primary divorce case records
Data period	2022 to June 2024
Initial number of records	2,424 divorce case records
Final unit of analysis	Subdistrict
Clustering method	K-Means Clustering
Number of clusters	3 clusters
Evaluation method	Davies-Bouldin Index

2.2. Data Selection and Preprocessing

The data selection stage was conducted to identify attributes relevant to the research objectives. The selected attributes were subdistrict, age difference between spouses, and factors causing divorce. The age difference was calculated from the absolute difference between the ages of the spouses and then categorized into small, moderate, and large age differences. Data cleaning was performed to address incomplete, inconsistent, or irrelevant records. Afterward, categorical data related to divorce-causing factors were transformed into numerical codes to enable processing using the K-Means algorithm. The data were then aggregated by subdistrict based on the frequency of age-difference categories and divorce-causing factors.



No	Jenis Perkara	Faktor Penyebab	Tgl AC	Pemohon	Kecamatan	Desa	Usia	Pekerjaan	Pendidikan	Termohon	Usia.1	Pekerjaan.1	Pendidikan.1
1	Cerai Gugat	Ekonomi	01/03/2022	NAMA P1	Dringu	Tamansari	55	Pedagang	SD	NAMA T1	57	Petani	SD
2	Cerai Gugat	Peselisihan dan Pertengkaran Terus Menerus	01/03/2022	NAMA P2	Pakuniran	Glagah	43	Buruh Tani / Perkebunan	SD	NAMA T2	42	Buruh Tani / Perkebunan	SLTA
3	Cerai Gugat	Ekonomi	01/03/2022	NAMA P3	Sukapura	Ngepung	24	Ibu Rumah Tangga	SLTA	NAMA T3	26	Petani	SLTA
4	Cerai Gugat	Peselisihan dan Pertengkaran Terus Menerus	01/03/2022	NAMA P4	Gading	Wangkal	25	Ibu Rumah Tangga	SLTA	NAMA T4	27	Buruh	SLTA
5	Cerai Gugat	Ekonomi	01/03/2022	NAMA P5	Gending	Curahsawo	27	Bidan	D3	NAMA T5	31	Perawat	D3

Figure 2. Data Selection, Cleaning, and Transformation Process

Table 2. Attributes Used in the Analysis

Attribute	Data Type	Processing Stage	Role
Subdistrict	Categorical	Used as regional aggregation identity	Cluster result identifier
Spouses' age	Numeric	Used to calculate age difference	Initial variable
Age difference	Numeric	Categorized into small, moderate, and large differences	Clustering feature
Divorce-causing factor	Categorical	Converted into numerical codes and counted by frequency	Clustering feature

The age-difference categories used in this study are presented in Table 3.

Table 3. Spousal Age Difference Categories

Category	Age Difference Range
Small age difference	Less than 5 years
Moderate age difference	5 to 10 years
Large age difference	More than 10 years

2.3. K-Means Clustering

K-Means was used to cluster subdistricts based on similarities in divorce-case characteristics. This study employed three clusters to form high, moderate, and low divorce-case profile groups. The K-Means process consisted of determining the number of clusters, selecting initial centroids, calculating the distance of data points to centroids, assigning cluster members, updating centroids, and repeating the process until the clusters became stable.

The distance between a data point and a centroid was calculated using Euclidean Distance as follows:

[
d(x_i,c_k)=\sqrt{\sum_{j=1}^{m}(x_{ij}-c_{kj})^2}
]

Where:
(d(x_i,c_k)) is the distance between data point (i) and centroid (k);
(x_{ij}) is the value of data point (i) on attribute (j);
(c_{kj}) is the value of centroid (k) on attribute (j);
(m) is the number of attributes.

After each data point was assigned to its nearest cluster, a new centroid was calculated using the average value of all members within that cluster. The iteration stopped when there was no change in cluster membership or when the centroid positions reached stability.

2.4. Evaluation and Streamlit Deployment

The clustering results were evaluated using the Davies-Bouldin Index. DBI measures the ratio between the spread of data within a cluster and the distance between clusters. A lower DBI value indicates better cluster quality because members within a cluster are more similar and the separation between clusters is clearer.

[
DBI=\frac{1}{K}\sum_{i=1}^{K}\max_{j\ne i}\left(\frac{S_i+S_j}{M_{ij}}\right)
]

Where:
(K) is the number of clusters;
(S_i) and (S_j) are the dispersion levels of data within clusters;
(M_{ij}) is the distance between the centroids of clusters (i) and (j).

2.5. Streamlit Implementation

The clustering model was implemented using Streamlit. The application provides features for dataset upload, initial data display, preprocessing results, transformed data, clustering process, cluster results, DBI value, and scatter plot visualization. This implementation aims to make it easier for users

to view the analysis results without directly running the code in a programming environment.

3. RESULTS

3.1. Data Preparation Results

The initial dataset consisted of 2,424 divorce case records. After selection and cleaning, the data were processed using the attributes of subdistrict, spouses' age difference, and divorce-causing factors. The age difference was calculated from the difference between the ages of petitioners and respondents. Divorce-causing factors were transformed from categorical values into numerical representations. The transformed data were then aggregated by subdistrict. Each subdistrict was represented by the frequency of cases in each age-difference category and divorce-causing factor. This process produced numerical subdistrict-level data that could be used as input for the K-Means algorithm.

Kecamatan	selisih dekat	selisih sedang	selisih jauh	Cacat Badan	Dihukum Penjara	Ekonomi	Judi	Kawin Paksa	Dalam Rumah Tangga	Mabuk	Salah Satu Pihak	Murtad	Pertengkaran Terus Menerus	Poligami	Zina
Dringu	71	20	13	1.0	1.0	45	0.0	2.0	1.0	3.0	8	0.0	50	0.0	1.0
Pakuniran	64	21	17	1.0	1.0	52	0.0	0.0	1.0	2.0	7	1.0	42	0.0	1.0
Sukapura	23	5	3	0.0	0.0	14	0.0	0.0	1.0	0.0	2	0.0	18	0.0	0.0
Gading	68	26	14	1.0	1.0	59	0.0	0.0	1.0	0.0	5	0.0	46	0.0	0.0
Gending	57	14	20	0.0	1.0	57	0.0	0.0	3.0	1.0	5	0.0	33	0.0	0.0
Pajarakan	40	21	7	0.0	0.0	41	0.0	0.0	1.0	0.0	5	0.0	27	1.0	0.0
Kuripan	23	9	7	0.0	0.0	20	0.0	0.0	2.0	0.0	2	0.0	16	0.0	0.0
Banyuanyar	82	35	12	0.0	2.0	78	0.0	0.0	3.0	1.0	7	0.0	53	1.0	1.0
Besuk	73	26	21	0.0	0.0	59	1.0	2.0	3.0	1.0	5	0.0	53	1.0	1.0
Maron	97	40	16	0.0	0.0	80	2.0	1.0	5.0	4.0	1	1.0	64	0.0	0.0
Tegalsiwalan	61	20	15	0.0	1.0	53	0.0	1.0	2.0	1.0	4	0.0	41	1.0	0.0
Wonomerto	40	19	14	0.0	0.0	31	0.0	1.0	1.0	0.0	2	0.0	43	1.0	1.0
Krucil	51	16	4	0.0	0.0	31	0.0	0.0	1.0	1.0	5	1.0	43	0.0	0.0
Paiton	89	26	12	0.0	1.0	64	0.0	0.0	3.0	1.0	6	0.0	61	1.0	0.0
Krejengan	51	30	17	0.0	0.0	51	1.0	1.0	1.0	0.0	4	0.0	47	0.0	1.0
Bantaran	31	12	7	0.0	0.0	26	0.0	0.0	2.0	0.0	2	0.0	28	0.0	0.0
Kraksaan	111	34	9	0.0	1.0	60	0.0	0.0	2.0	1.0	5	1.0	92	0.0	1.0
Kotaanyar	45	20	6	1.0	0.0	30	0.0	0.0	1.0	0.0	3	0.0	45	0.0	0.0
Sumberasih	70	34	11	0.0	0.0	53	1.0	1.0	3.0	1.0	3	0.0	62	0.0	1.0
Leces	72	37	16	1.0	1.0	66	1.0	2.0	3.0	1.0	5	0.0	54	0.0	1.0
Tiris	66	23	19	0.0	0.0	56	0.0	0.0	2.0	1.0	11	0.0	51	0.0	0.0
Tongas	90	21	12	0.0	0.0	65	1.0	0.0	3.0	1.0	6	0.0	54	0.0	0.0
Sumber	27	2	1	0.0	0.0	15	0.0	0.0	0.0	0.0	2	0.0	14	0.0	0.0

Figure 3. Divorce Case Data Transformation Results

3.2. Clustering Results

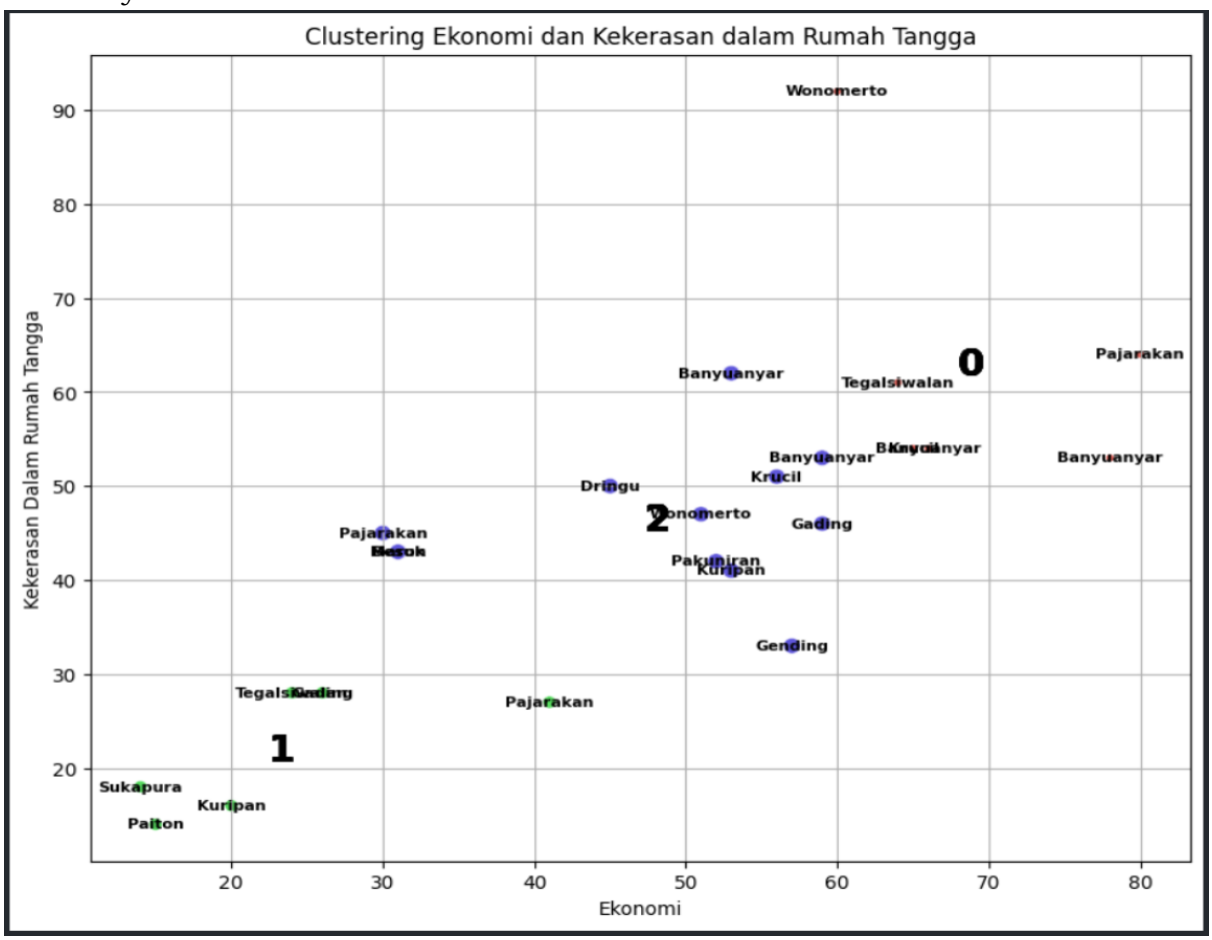
The K-Means process with three clusters produced groupings of subdistricts based on similarities in divorce-case characteristics. The results were divided into high, moderate, and low divorce-case profile clusters.

Table 4. Divorce Case Clustering Results

Cluster	Case Profile	Number of Subdistricts	Subdistricts
Cluster 0	High	1	Maron
Cluster 1	Low	9	Sukapura, Pajarakan, Kuripan, Wonomerto, Krucil, Bantaran, Kotaanyar, Sumber, and Lumbang
Cluster 2	Moderate	14	Dringu, Pakuniran, Gading, Gending, Banyuanyar, Besuk, Tegalsiwalan,

			Paiton, Krejengan, Kraksaan, Sumberasih, Leces, Tiris, and Tongas

Based on Table 4, Maron Subdistrict belongs to Cluster 0 as the group with the highest divorce-case profile. Cluster 1 consists of nine subdistricts with a low divorce-case profile. Meanwhile, Cluster 2 represents a moderate divorce-case profile and contains the largest number of subdistricts. Economic problems and continuous disputes or arguments were the factors that frequently appeared in divorce cases. Other factors with lower frequencies included abandonment of one spouse, domestic violence, forced marriage, underage marriage, imprisonment, and alcoholism. Physical disability, apostasy, gambling, adultery, and polygamy were the least frequent factors in the analyzed data.



Gambar 4. Visualisasi Hasil K-Means Clustering

3.3. Davies-Bouldin Index Evaluation

The internal evaluation of the clustering results using the Davies-Bouldin Index produced a value of 0.6611275396701337. A lower DBI value indicates that clusters are more compact internally and better separated from one another. With a value of 0.6611, the three-cluster result can be used to provide an initial overview of divorce-case grouping patterns by subdistrict. The DBI value should not be interpreted as classification accuracy. DBI is a measure of cluster structure quality because K-Means operates without predefined

labels. To determine the most optimal number of clusters, future testing should compare DBI values across several cluster alternatives, such as (k=2), (k=3), (k=4), and (k=5).

Table 5. Davies-Bouldin Index Evaluation Result

Number of Clusters Used	DBI Value	Interpretation
3	0.6611	Reasonably good cluster separation

3.4. Streamlit Application Results

The Streamlit application was used to present the clustering process through an interactive interface. On the initial page, users can upload a dataset for analysis. Once the data are uploaded, the application displays the dataset contents and preprocessing results.

Figure 5. Streamlit Application Home Page

The next stage displays transformed data and the clustering result table. Users can review the subdistricts included in each cluster along with the scatter plot visualization. The application also displays the Davies-Bouldin Index value as a measure of cluster quality.

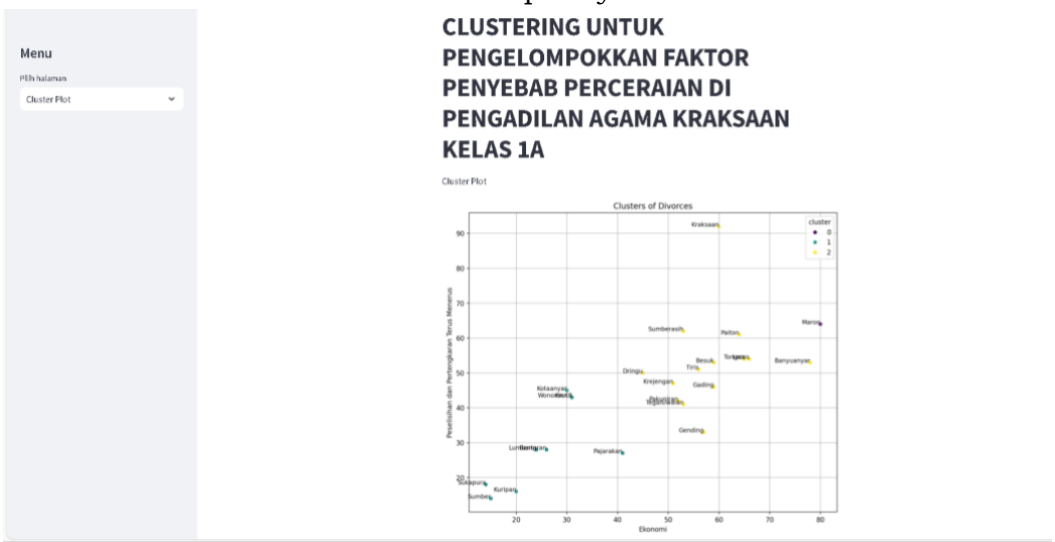


Figure 6. Clustering Results in the Streamlit Application

The Streamlit implementation shows that K-Means analysis results can be presented through a simple web application. The application can support data monitoring; however, final interpretation still needs to be performed by relevant administrators based on legal, social, and administrative contexts of divorce cases.

4. CONCLUSIONS

This study successfully applied the K-Means Clustering algorithm to group divorce cases at the Kraksaan Class 1A Religious Court. The research data consisted of 2,424 divorce case records from 2022 to June 2024. The variables used included subdistrict, spouses' age difference, and factors causing divorce. After data selection, data cleaning, transformation, and aggregation by subdistrict, the K-Means process formed three clusters. Cluster 0 represents the group with the highest divorce-case profile and consists of Maron Subdistrict. Cluster 1 represents the low divorce-case group and consists of nine subdistricts, while Cluster 2 represents the moderate divorce-case group and consists of fourteen subdistricts. Economic problems and continuous disputes or arguments were the most dominant factors found in the divorce case data. Evaluation using the Davies-Bouldin Index produced a value of 0.6611. This value indicates that the resulting cluster structure has reasonably good separation quality. However, DBI is not a classification accuracy measure; it is an internal clustering evaluation metric. The Streamlit implementation allows users to upload data, view preprocessing results, examine clustering outcomes, and review the analysis visualization more easily. Future research is recommended to compare K-Means with other methods, such as K-Medoids, Fuzzy C-Means, DBSCAN, or Spectral Clustering. Further studies may also include additional variables, such as case type, educational level, occupation, duration of marriage, number of children, and mediation history. Testing multiple cluster alternatives and applying feature normalization are also recommended to strengthen and improve the comparability of clustering results.

5. REFERENCES

Abu, T. (2023). Buku referensi implementasi algoritma machine learning berbasis web dengan framework Streamlit. Pustaka Nurja. https://pustakanurja.unuja.ac.id

Alvianta, F. N., Prabowo, A. A., & Komarudin, A. (2021). Pembinaan kesejahteraan keluarga dalam pemberdayaan keluarga prasejahtera. JISIP: Jurnal Ilmu Sosial dan Pendidikan, 5(3), 137–151. https://doi.org/10.36312/jisip.v5i3.2095

Ananda, D. P., & Monalisa, S. (2023). Segmentasi pelanggan B2B dengan model LRFM menggunakan algoritma Fuzzy C-Means pada Rotte Bakery. Jurnal Teknologi Informasi dan Ilmu Komputer, 10(5). https://doi.org/10.25126/jtiik.2023106569

Bezdek, J. C. (1981). Pattern recognition with fuzzy objective function algorithms. Plenum Press.

Herdiana, D. (2020). Pengawasan kolaboratif dalam pelaksanaan kebijakan bantuan sosial terdampak COVID-19. JDP: Jurnal Dinamika Pemerintahan, 3(2), 85–99. https://doi.org/10.36341/jdp.v3i2.1323

Huddin, S., Haerani, E., Jasril, J., & Oktavia, L. (2023). Penerapan Fuzzy C-Means pada klasterisasi penerima bantuan pangan non tunai. KLIK: Kajian Ilmiah Informatika dan Komputer, 4(1), 453–461. https://doi.org/10.30865/klik.v4i1.988

Prajetno, J., Setiawan, E. I., & Putra, A. S. (2022). Pembentukan aturan fuzzy untuk

pemberian rekomendasi penerima bantuan keluarga berumah tidak layak huni menggunakan K-Means clustering. Information System and Technology, 4(2), 85–92. https://doi.org/10.52985/insyst.v4i2.216

Rahmadina, E. S., Irawan, B., & Bahtiar, A. (2024). Penerapan algoritma Fuzzy C-Means pada data penjualan distro. JATI: Jurnal Mahasiswa Teknik Informatika, 8(2), 2348–2354. https://doi.org/10.36040/jati.v8i2.8467

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sudriyanto, S., Khairi, A., & Hikam, A. S. (2023). Penerapan algoritma K-Means untuk clustering santri pra-sejahtera di Yayasan Bantuan Sosial Az-Zainiyyah Pondok Pesantren Nurul Jadid. NJCA: Nusantara Journal of Computers and Its Applications, 8(1), 22–31.

Supriyanti, K. R., Damiri, B. A., & Ramadhan, W. N. (2024). Pengelompokan faktor-faktor yang mempengaruhi curah hujan di Provinsi Sumatera Utara menggunakan metode Fuzzy C-Means. Jurnal Statistika dan Komputasi, 3(1), 1–10. https://doi.org/10.32665/statkom.v3i1.2623