

Analysis of Age Patterns and Factors Causing Marriage Dispensation Using the K-Means Algorithm (Case Study: Stabat Religious Court)

Della Yulia Citra^{1*}, Achmad Fauzi², Ratih Puspadini³

^{1,2,3}STMIK Kaputama Binjai

dellayuliacitra8@gmail.com^{1*}, fauzyriva88@gmail.com², puspadini.ratih@gmail.com³

Abstract

The Stabat Religious Court has data on marriage dispensation applications that need to be analyzed to determine the age pattern, causal factors, and domicile of the applicants. This study aims to analyze the age pattern and causal factors of marriage dispensation and group the applicant data based on age, reason for application, and domicile using the K-Means algorithm. The data used were 749 marriage dispensation applications from 2021–2025. The study was conducted through the Knowledge Discovery in Database (KDD) stages, namely Data Selection, Data Preprocessing, Transformation, Data Mining, and Evaluation with the help of the MATLAB application. Determining the optimal number of clusters using the Elbow method based on the Sum of Squared Error (SSE) value resulted in three clusters. The clustering results showed that Cluster 1 had 210 data, Cluster 2 had 150 data, and Cluster 3 had 389 data. The analysis results showed that the age pattern of applicants who most frequently applied for marriage dispensation was 17 years old, while the most consistent causal factor was the reason for avoiding adultery. Cluster 3 has the largest number of data points, comprising 389 individuals. The dominant characteristics of applicants are 17 years old, the reason being to avoid adultery, and their domicile in Hinai District. Therefore, the application of the K-Means algorithm is able to group marriage dispensation data based on age, reason, and domicile, generating information that can be used to support data analysis and efforts to prevent early marriage at the Stabat Religious Court.

Keywords: Cluster, K-Means, Elbow, Matlab, Marriage_Dispensation, PA

1. Introduction

Several previous studies have discussed the issue of early marriage and the application of the K-Means method to group prospective bride and groom data based on certain characteristics so that data groups that have similarities can be identified. Research by Nst et al. (2023) discusses factors related to the occurrence of early marriage, such as economic conditions, education, and social factors. Other studies also show that clustering analysis can be used to identify certain patterns or characteristics of a data set. However, these studies have not specifically grouped marriage dispensation application data based on a combination of variables of applicant age, reason for application, and area of domicile at the Stabat Religious Court. Therefore, this study was conducted to fill this gap by applying the K-Means algorithm so that the characteristic patterns of marriage dispensation applicants can be identified in a more structured manner.

2. Research Methods

This study employs the KDD method to analyze 749 marriage dispensation applications filed at the Stabat Religious Court between 2021 and 2025. The KDD process comprises data selection, data preprocessing, transformation, data mining, and evaluation. The data were analyzed using three variables: applicant age, reason for the application, and region of domicile. Categorical variables specifically the reason for the application and domicile were transformed into numerical codes to enable processing by the K-Means algorithm. Subsequently, K-Means was used to group the data based on shared characteristics, and the resulting clusters were evaluated based on their centroids and specific group characteristics; the following figure illustrates the KDD method



Fig.1 Research Methodology Structure

2.1. Systems Analysis

Marriage dispensation data at the Stabat Class IB Religious Court was obtained from the Case Tracking Information System (SIPP) application which contains various types of cases, such as divorce, child custody, inheritance, marriage confirmation, and marriage dispensation. This condition causes the marriage dispensation data not to be specifically grouped so that the process of searching and analyzing information still requires quite a long time. Therefore, this study applies the K-Means Clustering method using MATLAB to group data based on the applicant's age, reason for application, and domicile area, so that the pattern and characteristics of marriage dispensation at the Stabat Religious Court can be identified.

2.2. Marriage Dispensation

According to research from (sebyar 2022) on marriage dispensation, it is a court permission for prospective husbands or wives under 19 years of age to marry, taking into account the best interests of the child. Previous research shows that early marriage is influenced by factors such as pregnancy, economics, education, and social environment, while *K-Means* can be used to group data based on its characteristics. However, research that groups marriage dispensation based on age, reason for application, and area of domicile is still limited. This study fills the gap by applying *K-Means* to marriage dispensation data at the Stabat Religious Court.

2.3. Problem Solving Methods

The research problem is addressed using the K-Means clustering algorithm based on the Knowledge Discovery in Databases (KDD) methodology. The KDD process consists of Data Selection, Data Preprocessing, Transformation, Data Mining, and Evaluation, which systematically transform raw data into meaningful information.

2.3.1 Data Mining

Data mining is the process of gathering valuable information from a large amount of data. The data mining process typically utilizes mathematical and statistical methods, as well as artificial intelligence technology. Therefore, data mining is the analytical process of reviewing data sets to discover unexpected patterns or relationships and summarize the data in a way that is understandable and useful to the data owner.

2.3.2 Clustering

Clustering is a data mining technique that groups data with similar characteristics into the same cluster. This study employs the K-Means algorithm because of its simplicity, computational efficiency, and effectiveness in clustering numerical data based on the minimum distance to cluster centroids.

2.3.3 MATLAB

MATLAB was used to normalize the data, determine the optimal number of clusters using the Elbow Method, perform K-Means clustering, and visualize the clustering results. The software also facilitated the analysis of cluster centroids and member distribution.

2.3.4 Research Supporting Data for 2025

The dataset used in this study was obtained from marriage dispensation applications at the Stabat Religious Court for the 2021–2025 period. A total of 749 marriage dispensation applications were analyzed. This study used three variables: the age of the applicant, the reason for the marriage dispensation application, and the applicant's domicile. This third variable was chosen because it represents the characteristics of marriage dispensation applications based on age, the reason for the application, and the distribution of the applicant's domicile. Table 1 presents the sample dataset used in this study.

Table 1. Sample Data For 2021-2025

No	Age	Reason	Area Of Domicile	Length of Relationship	Case Number
1	16 years	Avoiding Adultery	Jln. Makmur Dusun IV, Desa Banyumas, Kecamatan Stabat, Kabupaten Langkat	3 years	*/***.P/2021/PA.Stb
	13 years	Pregnancy	Jl. Sempurna, Kelurahan Sidomulyo, Kecamatan Stabat, Kabupaten Langkat	1 years	*/***.P/2021/PA.Stb
3	17 years	Avoiding Adultery	Lingkungan Perdamaian, Kelurahan Tangkahan Durian, Kecamatan Brandan Barat, Kabupaten Langkat	2 years	*/***.P/2022/PA.Stb
4	15 years	Pregnancy	Dusun II, Desa Bekulap, Kecamatan Selesai, Kabupaten Langkat,	2 years	*/***.P/2022/PA.Stb
5	18 years	Promiscuity	Dusun Dondong Timur II, Desa Stabat Lama Barat, Kecamatan Wampu, Kabupaten Langkat	2 years	*/***.P/2023/PA.Stb
6	17 years	Costoms/Culture	Dusun III, Desa Pematang Guntung, Kecamatan Teluk Mengkudu, Kabupaten Serdang Bedagai;	3 years	*/***.P/2023/PA.Stb
7	18 years	Avoiding Adultery	Dusun VIII Bandar Meriah, Desa Namu Ukur Utara, Kecamatan Sei Bingai, Kabupaten Langkat ;	2 years	*/***.P/2024/PA.Stb
8	17 years	Promiscuity	Jln. Pangkalan Brandan Gang Lurah Lingkungan V, Kelurahan Beras Basah, Kecamatan Pangkalan Susu, Kabupaten Langkat	1 years	*/***.P/2024/PA.Stb
9	16 years	Pregnancy	Lingkungan IV Paya Mabar Inpres, Kelurahan Paya Mabar, Kecamatan Stabat, Kabupaten Langkat	1 years	*/***.P/2025/PA.Stb
10	18 years	Avoiding Adultery	Cinta Dapat Dusun Mawar, Desa Padang Brahrang, Kecamatan Selesai, Kabupaten Langkat	2 years	*/***.P/2025/PA.Stb

3. Result and Discussion

3.1. Algoritma K-Means

The K-Means algorithm is a non-hierarchical clustering technique that divides data into one or more clusters. The K-Means algorithm groups data with similar characteristics into a single cluster by randomly determining the number of clusters (centroids) and calculating the distance between data using the Euclidean distance equation. The distance calculation between data points and centroids is performed using the Euclidean distance equation:

$$D_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

The centroid value is also updated by calculating the average of all data in the cluster using the Euclidean formula:

$$C(i) = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x}$$

3.2. Interface Discussion

Which displays the results of the elbow and a significant decrease in SSE values before reaching the third cluster. graph to produce the best k value which is seen at k = 3 by forming a very clear elbow and experiencing a significant decrease from the previous k value, so that when the value of k = 3, the results of this grouping and analysis are data with the best grouping.

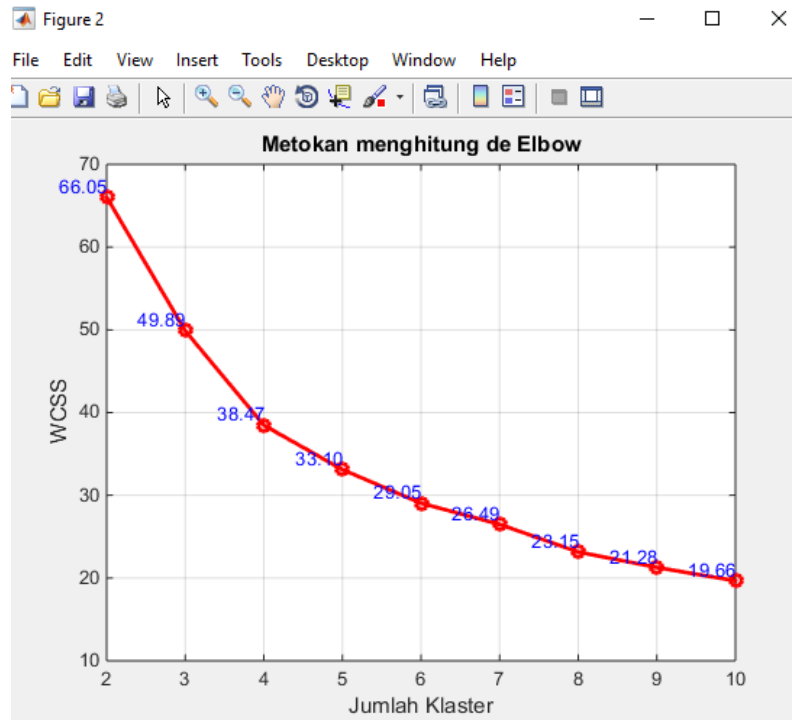


Fig 2. Optimal Elbow Cluster Results

3.3. Program Analysis

Program analysis was conducted to initiate the clustering process using MATLAB. The K-Means algorithm completed the clustering process after three iterations, yielding three optimal clusters. The results comprised 210 applications in Cluster 1, 150 in Cluster 2, and 389 in Cluster 3—representing the marriage dispensation applications at the Stabat Religious Court. The resulting clusters were then analyzed to determine the characteristics of each group, with interpretation based on centroid values.

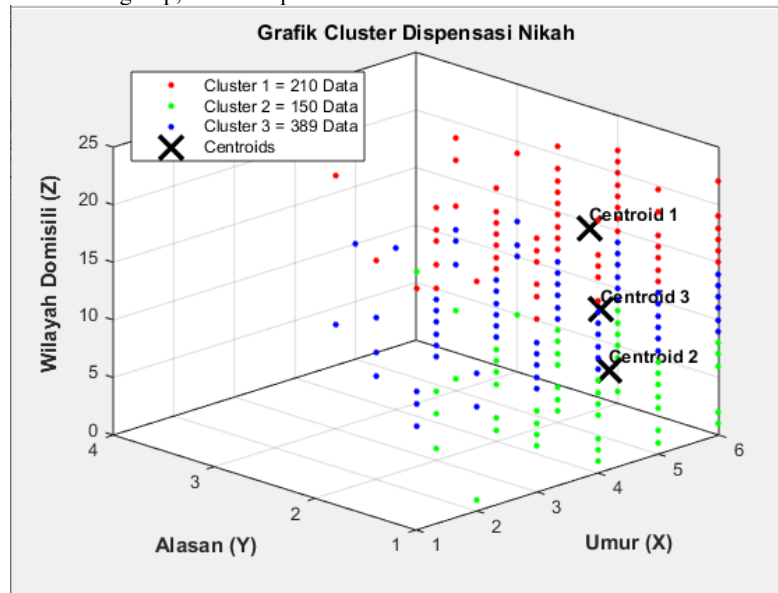


Fig 3. 3 Cluster Test Results

3.4. Cluster Interpretation

Based on the centroid analysis, Cluster 3 represents a group of marriage dispensation applicants characterized by younger ages, specific dominant reasons for the application, and residence in regions with higher application volumes. Cluster 1 represents a group with intermediate characteristics, while Cluster 2 represents a group with distinct characteristics regarding age, reasons for the application, and region of residence. These clustering results indicate that marriage dispensation applications can be categorized into groups sharing similar characteristics based on these three variables.

4. Results

1. Cluster 1 results that can be known the results of the Age (X), Reason (Y), Domicile Area (Z) group with the number of data owned as many as 210 data. From the data description shows that the applicant's age pattern with the centroid point on the age variable is 4.97 then the applicant's reason with the center point of 1.65 and the domicile area is 2.46. With the results of the information on Age (X) "17 years", Reason (Y) "Avoiding Adultery", Domicile Area (Z) "Babalan District".
2. The results of Cluster 2 which can be seen are the results for Age group (X), Reason (Y), Area of Domicile (Z) with a total of 150 data. From the data information, it shows that the applicant's age pattern with the centroid center point on the age variable is 5.10, then the applicant's reasons with the center point are 1.54 and the domicile area is 5.62. With the results of the information on Age (X) "17 years", Reason (Y) "Avoiding Adultery", Area of Domicile (Z) "Kuala District".
3. Cluster 3 results that can be known the results of the Age (X), Reason (Y), Domicile Area (Z) group with the number of data owned as many as 389 data. From the data description shows that the applicant's age pattern with the centroid point on the age variable is 4.95 then the applicant's reason with the center point of 1.54 and the domicile area is 11.10. With the results of the information on Age (X) "17 years", Reason (Y) "Avoiding Adultery", Domicile Area (Z) "Hinai District".

5. Conclusion

Based on the research results, the K-Means algorithm was successfully applied to analyze the age patterns, causal factors, and domicile areas of marriage dispensation applicants at the Stabat Religious Court using 749 data sets for the 2021–2025 period. The optimal number of clusters was determined using the Elbow method, then the clustering process was carried out using the Knowledge Discovery in Database (KDD) stage and implemented in MATLAB. The clustering results showed that 17 years old was the most dominant age group applying for marriage dispensation, while the most frequently occurring causal factors were pregnancy, avoiding adultery, promiscuity, and customs/culture. The clustering resulted in Cluster 1 with 210 data sets, Cluster 2 with 150 data sets, and Cluster 3 with 389 data sets. These results are able to provide information regarding the characteristics of marriage dispensation applicants so that they can be supporting materials for the Stabat Religious Court in analyzing data and supporting efforts to prevent early marriage.

References

- [1] Arifiani Rina, Rito Cipta Sigitta Hariyono, Haries Anom Susetyo Aji Nugroho, Sri Hartati et al. 2025. "Pemanfaatan Data Mining Dengan K-Means Clustering Dalam Penilaian Guru Berprestasi." 9(3): 1124–38..
- [2] Dwi, Vicco, Arya Putra, and Raditia et al. Vindua. 2025. "Implementasi Data Mining Dalam Optimal Stok Bahan Baku Kue Di Pt. XYZ Menggunakan Metode K-Nearst Neighbor." 9(3): 4397–4403.
- [3] Fatwa, Muhamad, Rizki Ristu, Sriwinarty Pandiangan, and Edi Supriyadi. 2022. "Pengaplikasian Matlab Pada Perhitungan Matriks." 1: 81–93.
- [4] Febrianti, Tia, and Erwin Harahap. 2021. "Penggunaan Aplikasi MATLAB Dalam Pembelajaran Program Linear The Use of MATLAB Applications in Linear Programming Learning." 20(1): 1–7.
- [5] Hanif, Naufal, Ratih Puspadini, Tangki Kapal, and Metodologi Penelitian. 2025. "Machine Translated by Google Prototipe Sistem Deteksi Kebocoran Tangki Kapal Berbasis IoT Untuk Meningkatkan Keamanan Machine Translated by Google." 5(1): 1–6.
- [6] Hasibuan, Bunga Lestari, Ega Yolanda, Universitas Islam, Negeri Sumatera, and Sumatera Utara. 2022. "Pengklasifikasian Data Pasien Tes Urine Dengan Metode Clustering Pada Kantor Badan Narkotika Nasional Provinsi Sumut (BNNP SUMUT)." 4(November): 183–93.
- [7] Herdawati, Ridwan Muannif. 2022. "Jurnal Indragiri." 2(2): 132–39.
- [8] Maori, Nadia Annisa. 2023. "Metode Elbow Dalam Optimasi Jumlah Clustering Pada K-Means." 14(2): 277–87.
- [9] Nst, Aisyah Amalia, Aknes Dini, Arisah Faslon, Tri Sunarsi, Dechoni Rahmawati, Universitas Jenderal, and Achmad Yani. 2023. "Dampak Pernikahan Dini Terhadap Kesehatan Reproduksi Literature Review(The Impact Of Early Marriage On Reproductive Health : A Literature Review)." 9(2): 126–33.
- [10] Nugraha, Agung, Odi Nurdian, Gifthera Dwilestari, Program Studi, Teknik Informatika, Kabupaten Cirebon, and Toko Olahraga. 2022. "Penerapan Data Mining Metode K-Means Clustering Untuk." 6(2): 849–55.
- [11] Nuraeni Septiani, Saeful Anwar, Ruli Herdiana. 2023. "Penerapan Algoritma K-Means Clustering Untuk Harga Rumah Di Jakarta Selatan." 1(2)74.
- [12] Nurlaila, Relita Buaton, Rusmin Saragih. 2022. "Data Mining Pengelompokan Akta Nikah Berdasarkan Usia Nikah Atau Domisili Menggunakan Metode Clustering: Studi Kasus Kemenag Langkat." 1: 1–11.
- [13] Nursaniah, Rini, Nining Rahaningsih, Irfan Ali, and Nisa Dienwati. 2024. "Pengelompokan Data Penerimaan Pajak Bumi Dan Bangunan Berdasarkan Kelurahan Dikota Tasikmalaya Menggunakan Algoritma K-Means ." 8(2): 1477–83.
- [14] Pratiwi, Sella Dwi, Achmad Fauzi, and I Gusti Prahmana. 2023. "Jurnal Kecerdasan Buatan Dan Aplikasi Rekayasa Pengelompokan Jumlah Anggota Perpustakaan Untuk Menentukan Lokasi Sosialisasi Menggunakan Metode Pengelompokan Machine Translated by Google." 3(1).
- [15] Rahmawati, Erika, Betha Nurina Sari, Mohamad Jajuli, Universitas Singaperbangsa Karawang, Telukjambe Timur, Mitigasi Bencana, and Jawa Tengah. 2024. "Implementasi Algoritma K-Means Pada Pemetaan Daerah." 8(3): 2568–74.
- [16] Safitri, Putri Chusnul Nur Iman. 2025. "Analisis Dispensasi Pernikahan Terhadap Jumlah Pernikahan." 2: 471–87.
- [17] Sari, Herlina Latipa. 2023. "Penerapan Data Mining Untuk Mengetahui Minat Siswa Pada Pelajaran IPA Menggunakan Metode K-Means Clustering." 2(4): 693–702.
- [18] Sebyar, Muhamad Hasan. 2022. "SYARI ' AH." 5(1): 1–14.
- [19] Sembiring, Julia Br, Hotler Manurung, and Anton Sihombing. 2023. "Pengelompokan Data Tunggalan Pembayaran Kredit Mobil Menggunakan Metode Clustering (Studi Kasus : CV Citra Kencana Mobil)." 3(July): 275–91.
- [20] Setiyo, Lorensius Anang, Ign. F. Bayu Andoro. 2021. "Industrial Engineering Conference 2021."
- [21] Siswanto, Cecep Tedi. 2025. "Analisis Kebijakan Dispensasi Kawin Di Indonesia." 5(1): 91–106.
- [22] Sulistiyawati, and Eko Supriyanto. 2021. "Implementasi Algoritma K-Means Clustering Dalam Penentuan Siswa Kelas Unggulan." 15(2): 25–36.
- [23] Zainab Tuasamu, Nur Afni Intan M.Lewaru, Muhammad Rivaldi Idris, Abdullah Bill NaZari Syafaat, Fitria Faradilla, Mariam Fadlan, Putri Nadiva, Rahmi Efendi. 2023. "Analisis Sistem Informasi Akuntansi Siklus Pendapatan Menggunakan DFD Dan Flowchart Pada Bisnis Porobico." 1(2).