

Penerapan Metode *K-Means* Untuk Pengelompokan Data Pelaporan Pernikahan

Nur Hayatin Nufus*, Sriani

Fakultas Sains dan Teknologi, Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}nufusnurhayatin7@gmail.com, ²sriani@uinsu.ac.id

Email Penulis Korespondensi: nufusnurhayatin7@gmail.com

Submitted: 20/11/2024; Accepted: 29/12/2024; Published: 30/12/2024

Abstrak—Proses pelaporan perkawinan di Kantor Urusan Agama (KUA) Kecamatan Laubaleng masih dilakukan secara manual, yang mengakibatkan kendala dalam pengelolaan, akurasi, dan efisiensi akses data. Penelitian ini bertujuan untuk mengatasi permasalahan tersebut melalui pengembangan sistem berbasis metode *K-Means Clustering*. Sistem ini dirancang untuk mengelompokkan data pelaporan perkawinan berdasarkan atribut seperti usia, status pernikahan, dan bulan pernikahan, sehingga dapat memberikan pola data yang lebih terstruktur dan informatif. Metode *Elbow* digunakan untuk menentukan jumlah kluster optimal, sedangkan algoritma *K-Means* diterapkan dengan menggunakan jarak *euclidean* untuk menghitung kedekatan data terhadap centroid. Proses penelitian melibatkan pengumpulan data pelaporan dari tahun 2019 hingga 2024, *preprocessing* data, normalisasi, serta evaluasi hasil *clustering* menggunakan *Davies-Bouldin Index* (DBI). Hasil penelitian menunjukkan bahwa metode *K-Means* efektif dalam mengelompokkan data, memberikan visualisasi yang jelas mengenai distribusi pola pernikahan, dan meningkatkan efisiensi pengelolaan data di KUA. Dengan sistem ini, KUA dapat meningkatkan kecepatan akses, mengurangi potensi kesalahan, dan mendukung pengambilan keputusan berbasis data yang lebih akurat.

Kata Kunci: K-Means; Pengelompokan Data; Pelaporan Perkawinan; Kantor Urusan Agama; Algoritma *Clustering*; *Davies-Bouldin Index*

Abstract—The process of reporting marriages at the Religious Affairs Office (KUA) in Laubaleng District is still done manually, which results in obstacles in the management, accuracy and efficiency of data access. This research aims to overcome this problem through developing a system based on the K-Means clustering method. This system is designed to group marriage reporting data based on attributes such as age, marital status, and month of marriage, so as to provide a more structured and informative data pattern. The Elbow method is used to determine the optimal number of clusters, while the K-Means algorithm is applied using Euclidean distance to calculate the closeness of the data to the centroid. The research process involves collecting reporting data from 2019 to 2024, data preprocessing, normalization, and evaluating clustering results using the Davies-Bouldin Index (DBI). The research results show that the K-Means method is effective in grouping data, providing clear visualization of the distribution of marriage patterns, and increasing the efficiency of data management at KUA. With this system, KUA can increase access speed, reduce the potential for errors, and support more accurate data-based decision making.

Keywords: K-Means; Data Mining; Marriage Reporting; Office of Religious Affairs; Clustering Algorithm; Davies-Bouldin Index

1. PENDAHULUAN

Pernikahan dalam Islam dianggap sebagai ibadah yang sangat penting, karena mengikat dua individu dalam ikatan yang diakui oleh Allah SWT [1]. Pernikahan dalam Islam dipandang sebagai ibadah yang mendatangkan keberkahan, karena melibatkan komitmen untuk hidup bersama dalam ketaatan kepada Allah SWT. Pernikahan juga dianggap sebagai cara untuk menjaga diri dari perbuatan zina (perzinahan) dan untuk melanjutkan keturunan. Pernikahan di Indonesia diatur oleh Undang-Undang Nomor 1 Tahun 1974 tentang Perkawinan, yang menetapkan bahwa pernikahan harus dilakukan dengan prosedur resmi dan diakui secara hukum. KUA bertanggung jawab untuk memfasilitasi proses ini sesuai dengan hukum yang berlaku [2]. Melalui KUA, masyarakat diberdayakan untuk memahami proses pernikahan yang sesuai dengan ajaran Islam. KUA juga memberikan penyuluhan dan pembinaan terkait pernikahan dan keluarga yang islami.

Perkawinan di KUA mencerminkan upaya pemerintah dalam mendukung pelaksanaan pernikahan yang sah, berkeadilan, dan sesuai dengan ajaran agama Islam. Dengan demikian, KUA diharapkan dapat menjadi fasilitator yang efektif dalam membangun keluarga yang *sakinah, mawaddah, warahmah* [3]. Saat ini proses pelaporan perkawinan di KUA Kecamatan Laubaleng masih dilakukan secara manual dan menggunakan sistem yang kurang efisien. Hal ini dapat menyebabkan kesulitan dalam mengakses dan memelihara data pelaporan perkawinan secara efektif. Terdapat kendala-kendala spesifik yang dihadapi oleh KUA Kecamatan Laubaleng dalam proses pelaporan perkawinan, baik dari segi teknologi maupun proses administrasi. Ini bisa menjadi masalah signifikan karena akses yang sulit dapat menghambat efisiensi dan akurasi dalam pencatatan dan pemeliharaan data tersebut. Proses pelaporan perkawinan yang belum optimal, menyebabkan potensi kesalahan atau keterlambatan dalam pengelolaan data. Hal ini mungkin disebabkan oleh metode manual atau sistem yang belum terotomatisasi dengan baik dan kurangnya *platform* yang memadai untuk memfasilitasi pelaporan dan pemeliharaan data dengan baik. Adanya tantangan dalam mengakses dan memelihara data pelaporan perkawinan di Kantor Urusan Agama (KUA) Kecamatan Laubaleng Kabupaten Karo ini bisa termasuk masalah aksesibilitas, pengelolaan manual yang memakan waktu, atau kesulitan dalam memastikan keakuratan data.

Proses pelaporan perkawinan tidak efisien, memakan waktu, dan rentan terhadap kesalahan manusia yang mana hal ini dapat mempengaruhi kinerja administratif dan pelayanan kepada masyarakat. Kesulitan dalam mengakses data

pelaporan perkawinan berdampak baik bagi petugas KUA maupun masyarakat yang membutuhkan informasi terkait [4][5]. Hal ini bisa menjadi hambatan dalam pengelolaan data yang efisien dan menyebabkan keterlambatan atau kebingungan dalam mendapatkan informasi yang diperlukan. Ketidakefektifan yang terjadi dalam pengelolaan dan pemanfaatan data pelaporan perkawinan di KUA Kecamatan Laubaleng ini dapat berdampak pada pengambilan keputusan yang kurang efektif atau kurang tepat berdasarkan data yang tersedia.

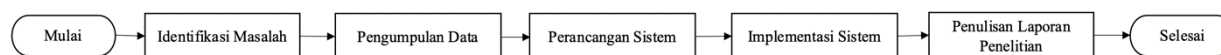
Adanya kebutuhan untuk memanfaatkan teknologi, seperti aplikasi *mobile*, untuk meningkatkan efisiensi dalam pengelolaan dan akses data pelaporan perkawinan. Penggunaan metode K-Means juga menunjukkan adanya upaya untuk menerapkan analisis data dalam meningkatkan pengelolaan informasi di KUA. Pengembangan aplikasi *mobile* ini diharapkan dapat membantu mengatasi beberapa kendala tersebut [6]. Dengan pengembangan aplikasi *mobile* dengan metode *K-Means* dapat dianggap sebagai upaya untuk meningkatkan efisiensi, aksesibilitas dan pemeliharaan data pelaporan perkawinan di KUA tersebut. Penelitian ini dilatar belakangi oleh *research GAP* pada penelitian-penelitian terdahulu. Berdasarkan penelitian yang dilakukan oleh Jatipaningrum, dkk [7] yang berjudul “Pengelompokan Kabupaten dan Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kesejahteraan Dengan Metode *K-Means* dan *Density-Based Spatial Clustering Of Applications With Noise*”. Penelitian ini berisi tentang pengelompokan kabupaten dan kota khususnya di Provinsi Jawa Timur berdasarkan tingkat kesejahteraan perlu dilakukan untuk mengetahui pemerataan tingkat kesejahteraan yang sudah dilakukan di Provinsi Jawa Timur. Oleh karena itu, untuk mengetahui pengelompokan kabupaten dan kota di Provinsi Jawa Timur berdasarkan kesamaan karakteristik tingkat kesejahteraannya kedalam sejumlah kelompok yang heterogen maka akan digunakan metode *K-Means* dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN). Perbandingan kedua metode menggunakan dua metode jarak yaitu *Euclidean* dan *Manhattan*, yang dibandingkan berdasarkan nilai *Davies-Bouldin Index* (DBI), *Hubert & Levin C-Index*, dan *Dunn Index* untuk dipilih metode terbaik.

Selanjutnya penelitian oleh Ramadhana, dkk [8] yang berjudul “Penerapan Data Mining Menggunakan Metode *K-Means Clustering* Pada Data Ekspor Batubara”. Penelitian ini menggambarkan Indonesia merupakan salah satu negara yang memiliki sumber daya alam batubara yang melimpah. Batubara banyak dibutuhkan berbagai negara seluruh dunia termasuk Indonesia untuk digunakan sebagai bahan utama pembangkit listrik. Berdasarkan data dalam sepuluh tahun terakhir, pembangkit listrik tetap menjadi konsumen batubara terbesar di Indonesia. Perusahaan batubara di Indonesia sejak tahun 2005 merupakan salah satu produsen dan eksportir batubara terbesar di dunia, melampaui produksi Australia yang ditunjukkan oleh ekspor batubara dari Indonesia. Data mining digunakan untuk membantu perusahaan mendapatkan informasi yang mereka butuhkan berdasarkan data yang ada. Penerapan data mining dalam hal ekspor batubara dapat dilakukan dengan membagi data tahun 2012 sampai dengan tahun 2021 menjadi beberapa kategori atau cluster. Hasil setiap kelompok yang berisi data yang sama dapat dianalisis pada kelompok tinggi, sedang dan rendah. *KDD (Knowledge Discovery in Database)* adalah proses komputasi yang menggunakan algoritma matematika yang dirancang untuk meng ekstrak data dan menghitung probabilitas kemungkinan tindakan dimasa depan [9]. Perbedaan penelitian-penelitian terdahulu dengan penelitian yang penulis lakukan yaitu terletak pada konteks penelitian yang berbeda. Setiap penelitian memiliki konteks berbeda, ada yang menggunakan *K-Means Clustering* dalam data ekspor batubara, pengelompokan kabupaten, dan administrasi Kantor Urusan Agama (KUA) seperti penelitian yang penulis lakukan. Penelitian yang penulis lakukan berkaitan dengan efektivitas *K-Means* dalam mengelompokkan data terkait Kantor Urusan Agama (KUA) sehingga memudahkan akses dan pemeliharaan data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah pengembangan dari kerangka penelitian yang akan menentukan keberhasilan dari suatu sistem yang dilakukan secara terencana, teratur, dan sistematis. Oleh karena itu, dalam penelitian ini, setiap tahapan dirancang dengan seksama untuk memastikan bahwa sistem ini dapat memenuhi kebutuhan pengguna dengan optimal. Berikut adalah tahapan-tahapan yang dilakukan dalam proses ini:



Gambar 1. Tahapan Penelitian

a. Identifikasi masalah

Tahap ini merupakan langkah awal dalam penelitian untuk menemukan dan merumuskan masalah yang akan diselesaikan. Fokus pada identifikasi masalah ini adalah memahami kebutuhan utama yang belum terpenuhi oleh sistem atau metode yang ada. Hasil dari tahap ini adalah perumusan masalah yang jelas dan spesifik sebagai dasar pengembangan sistem.

b. Pengumpulan Data

Pengumpulan data bertujuan untuk memperoleh informasi yang relevan dari populasi atau sumber data yang akan dianalisis. Data yang dikumpulkan harus memenuhi standar kualitas, seperti validitas, kelengkapan, dan relevansi, terutama jika metode seperti *K-Means* digunakan dalam proses penelitian.



c. Perancangan Sistem

Tahap ini bertujuan untuk merancang kerangka sistem yang akan dibangun, termasuk struktur *input*, *output*, dan proses yang dilakukan oleh sistem. Dalam penelitian berbasis *K-Means*, desain sistem harus mencakup integrasi algoritma *clustering* dan metrik jarak yang dipilih (misalnya, *Manhattan Distance*).

d. Implementasi Sistem

Tahapan ini melibatkan pengembangan sistem berdasarkan perancangan yang telah dibuat. Dalam hal ini, sistem akan dijalankan berdasarkan hasil rancangan sebelumnya. Sistem berbasis web dan algoritma *K-Means* diimplementasikan untuk menjalankan fungsi utama sistem.

e. Penulisan Laporan Penelitian

Penulisan laporan adalah proses mendokumentasikan seluruh tahapan penelitian, mulai dari identifikasi masalah hingga implementasi sistem. Laporan harus mencakup data yang dikumpulkan, hasil analisis, serta evaluasi sistem yang telah dibangun.

2.2 K-Means

Metode *K-Means* merupakan algoritma klusterisasi yang paling tua dan banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya [10]. Ide dasar algoritma *K-Means* sangatlah sederhana, yaitu meminimalkan *Sym of Squared Error* (SSE) antara objek–objek data dengan empat langkah [11].

Algoritma *K-Means Clustering* merupakan salah satu metode yang mengelompokkan data ke dalam kelompok–kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok dan data dengan karakteristik berbeda akan di kelompokkan ke dalam kelompok lain [12] [13].

Tujuan dari algoritma *K-Means* adalah mengelompokkan data ke dalam satu cluster dan meminimalkan kemiripan pada cluster lain. Ukuran kemiripan yang digunakan dalam *cluster* adalah fungsi jarak. Sehingga kemiripan data didapat dari jarak terpendek antar data terhadap titik pusat (*centeroid*) [14]. Tahapan algoritma *K-Means* adalah sebagai berikut [15]:

- a. Tentukan (*k*) sebagai jumlah cluster yang akan dibentuk.
- b. Tentukan titik pusat (*cluster*) awal acak.
- c. Hitung jarak dari setiap objek ke masing-masing *centroid* dari masing-masing cluster dengan jarak Euclidian, sebagaimana tertera pada persamaan (1).

$$d(x, y) = |x, y|$$

$$= \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3 \dots \dots, n \tag{1}$$

Keterangan: x_i adalah objek *x* ke-*i*, adalah daya *y* ke-*i* dan *n* adalah banyaknya objek

Karena pada penelitian ini menggunakan jarak Manhattan maka persamaan yang digunakan adalah persamaan (2).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \tag{2}$$

Keterangan: *d* adalah jarak antar *x* dan *y*, *x* adalah data pusat cluster dan *y* adalah data pada atribut.

- d. Kelompokkan masing-masing objek ke dalam *centroid* terdekat.
- e. Lakukan iterasi, kemudian hitung pusat cluster dengan persamaan (3)

$$v = \frac{\sum_{i=1}^n x_i}{n}; 1, 2, 3, \dots n \tag{3}$$

- f. Ulangi Langkah 3-4 hingga data tidak lagi berpindah ke cluster yang lain.

3. HASIL DAN PEMBAHASAN

3.1 Representasi Data

Penelitian ini menggunakan data yang dikumpulkan melalui observasi langsung di Kantor Urusan Agama (KUA) Kecamatan Laubaleng, studi pustaka, dan studi literatur. Data ini mencakup 225 catatan perkawinan dari tahun 2019 hingga 2024, dengan informasi seperti inisial nama pasangan, usia, dan status (lajang atau duda/janda). Data tersebut diolah menggunakan metode *K-Means* untuk mengidentifikasi pola-pola perkawinan, seperti usia rata-rata pasangan, waktu pelaksanaan akad, dan keterlibatan wali. Hasil pengelompokan ini memberikan wawasan bagi pihak KUA untuk menganalisis pola perkawinan, yang dapat digunakan sebagai dasar dalam merumuskan strategi pelayanan yang lebih efektif. Tabel 1 menyajikan ringkasan data perkawinan yang dikumpulkan selama periode penelitian.

Tabel 1. Distribusi Data Perkawinan 5 Tahun Terakhir

| No | Inisial Pasangan | Umur Calon Suami | Umur Calon Istri | Status Calon Suami | Status Calon Istri |
|----|------------------|------------------|------------------|--------------------|--------------------|
| | | | | | |



| | | | | | |
|-----|------------|-----|-----|--------|---------|
| 1 | JFS & RBM | 28 | 26 | Jejaka | Perawan |
| 2 | FS & EP | 23 | 20 | Jejaka | Perawan |
| 3 | M & SBS | 28 | 31 | Jejaka | Perawan |
| 4 | BG & INBK | 30 | 26 | Jejaka | Perawan |
| 5 | PAPA & TBT | 25 | 24 | Jejaka | Perawan |
| 6 | DS & NBP | 25 | 22 | Jejaka | Perawan |
| 7 | A & ST | 22 | 27 | Jejaka | Perawan |
| 8 | ST & ABG | 28 | 19 | Jejaka | Perawan |
| 9 | IKK & EYBG | 28 | 24 | Jejaka | Perawan |
| 10 | AS & ABK | 26 | 21 | Jejaka | Perawan |
| ... | ... | ... | ... | ... | ... |
| 25 | TS & KBBM | 26 | 24 | Jejaka | Perawan |

3.2 Transformasi Data

Transformasi data ini mencakup pengelompokan serta standarisasi nilai-nilai yang ada, sehingga memudahkan penerapan metode *K-Means* dalam tahap analisis berikutnya [16].

Tabel 2. Status Calon Suami

| Kode | Status Calon Suami |
|------|--------------------|
| 1 | Perjaka |
| 2 | Duda |

Tabel 3. Status Calon Istri

| Kode | Status Calon Istri |
|------|--------------------|
| 1 | Perawan |
| 2 | Janda |

3.3 Normalisasi Data

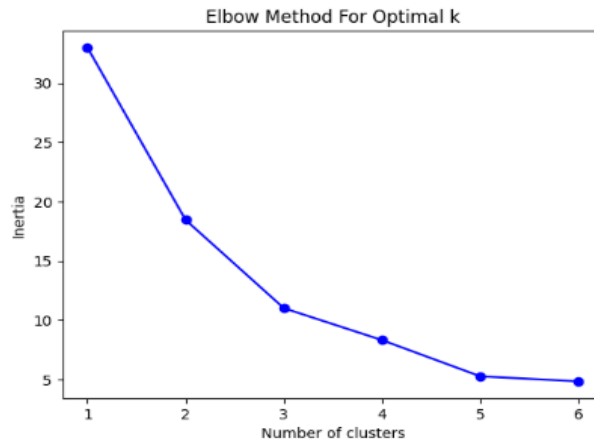
Normalisasi ini penting untuk menyamakan skala nilai pada setiap atribut, sehingga metode *K-Means* dapat menghasilkan pengelompokan yang lebih akurat dan seimbang [17]. Dengan normalisasi, data menjadi lebih konsisten dan siap untuk diproses dalam tahap analisis berikutnya, memastikan bahwa hasil pengelompokan mencerminkan pola-pola data secara optimal [18].

Tabel 4. Normalisasi Data

| No | Inisial Pasangan | Umur Calon Suami | Umur Calon Istri | Status Calon Suami | Status Calon Istri |
|-----|------------------|------------------|------------------|--------------------|--------------------|
| 1 | JFS & RBM | 0,23404 | 0,27778 | 1 | 1 |
| 2 | FS & EP | 0,12766 | 0,11111 | 1 | 1 |
| 3 | M & SBS | 0,23404 | 0,41667 | 1 | 1 |
| 4 | BG & INBK | 0,27660 | 0,27778 | 1 | 1 |
| 5 | PAPA & TBT | 0,17021 | 0,22222 | 1 | 1 |
| 6 | DS & NBP | 0,17021 | 0,16667 | 1 | 1 |
| 7 | A & ST | 0,10638 | 0,30556 | 1 | 1 |
| 8 | ST & ABG | 0,23404 | 0,08333 | 1 | 1 |
| 9 | IKK & EYBG | 0,23404 | 0,22222 | 1 | 1 |
| 10 | AS & ABK | 0,19149 | 0,13889 | 1 | 1 |
| ... | ... | ... | ... | ... | ... |
| 25 | TS & KBBM | 0,19149 | 0,22222 | 1 | 1 |

3.4 Metode Elbow

Metode *elbow* adalah teknik yang digunakan dalam *clustering* (khususnya pada algoritma *K-Means*) untuk menentukan jumlah *cluster* optimal dalam sebuah dataset. Fungsinya adalah untuk menemukan titik optimal di mana menambahkan lebih banyak cluster tidak memberikan perbaikan yang signifikan terhadap total *within-cluster sum of squares* (WCSS) atau variasi dalam cluster.



Gambar 2. Metode Elbow

Berdasarkan gambar 2, terlihat bahwa penurunan grafik mulai melambat setelah k = 3, yang menunjukkan adanya titik "elbow." Hal ini mengindikasikan bahwa jumlah kluster optimal untuk data ini adalah 3. Dengan memilih k = 3, kita dapat mencapai keseimbangan antara akurasi pengelompokan dan kompleksitas model, sehingga hasil analisis lebih efisien dan tetap representatif terhadap pola data.

3.5 Penerapan Metode K-Means

Setelah jumlah kluster ditentukan, langkah berikutnya adalah menetapkan centroid awal secara acak untuk inisialisasi proses *clustering* [19]. Dalam kasus ini, centroid awal dipilih dari hasil normalisasi data, di mana C0 diambil dari data ke-1, C1 dari data ke-14, dan C2 dari data ke-42. Pemilihan centroid ini penting untuk memulai algoritma K-Means dengan posisi awal yang baik, sehingga proses iterasi dapat menghasilkan pengelompokan yang optimal. Tabel di bawah ini menampilkan nilai centroid awal yang dipilih untuk setiap kluster [20]. Dapat dilihat pada tabel di bawah ini.

Tabel 5. Centroid Awal

| Centroid Awal | Umur Calon Suami | Umur Calon Istri | Status Calon Suami | Status Calon Istri |
|---------------|------------------|------------------|--------------------|--------------------|
| Centroid 0 | 0,23404 | 0,27778 | 0 | 0 |
| Centroid 1 | 0,27660 | 0,36111 | 1 | 0 |
| Centroid 2 | 0,29787 | 0,27778 | 1 | 1 |

Iterasi 1

Tabel 6. Hasil Iterasi 1

| Data | Jarak Centroid Iterasi 1 | | | Cluster |
|------|--------------------------|-------------|-------------|---------|
| | C0 | C1 | C2 | |
| 1 | 0,000000000 | 1,004368081 | 1,415653301 | 0 |
| 2 | 0,197726056 | 1,041481216 | 1,434137487 | 0 |
| 3 | 0,138890000 | 1,002446142 | 1,422450246 | 0 |
| 4 | 0,042560000 | 1,003465938 | 1,414373505 | 0 |
| 5 | 0,084623770 | 1,015189275 | 1,421050312 | 0 |
| 6 | 0,128139381 | 1,024268395 | 1,424304219 | 0 |
| 7 | 0,130647633 | 1,015903859 | 1,427389277 | 0 |
| 8 | 0,194450000 | 1,038736291 | 1,428945440 | 0 |
| 9 | 0,055560000 | 1,010495812 | 1,416743160 | 0 |
| 10 | 0,145261607 | 1,027922877 | 1,424993732 | 0 |
| ... | ... | ... | ... | ... |
| 25 | 0,069981541 | 1,013180213 | 1,419296875 | 0 |

Setelah perhitungan menggunakan *Euclidean Distance* dilakukan, data dikelompokkan berdasarkan jarak minimum ke centroid terdekat.

cluster lama (0 0)

cluster baru (0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0)

Terjadi perubahan pengelompokan data maka akan dilanjutkan ke perhitungan iterasi berikutnya.

Sebelum melanjutkan ke iterasi ke-2, peneliti perlu menentukan centroid baru dengan menghitung nilai rata-rata setiap atribut pada data yang termasuk dalam cluster 0, cluster 1 dan cluster 2.



Adapun centroid baru yang di peroleh adalah:

$$C0 = (0,195339762; 0,230423143; 0; 0,019047619)$$

$$C1 = (0,4521275; 0,4340275; 1; 0)$$

$$C2 = (0,56535; 0,722221429; 0,571428571; 1)$$

Iterasi 2

Tabel 7. Hasil Iterasi 2

| Data | Jarak Centroid Itrasi 2 | | | Cluster |
|------|-------------------------|-------------|-------------|---------|
| | C0 | C1 | C2 | |
| 1 | 0,064056164 | 1,035362467 | 1,310572238 | 0 |
| 2 | 0,138488223 | 1,09979765 | 1,387468744 | 0 |
| 3 | 0,19117639 | 1,023652011 | 1,286592524 | 0 |
| 4 | 0,095962024 | 1,027240568 | 1,300465239 | 0 |
| 5 | 0,032582331 | 1,060348949 | 1,341523775 | 0 |
| 6 | 0,071125101 | 1,072826878 | 1,357044182 | 0 |
| 7 | 0,117992366 | 1,065854227 | 1,34273726 | 0 |
| 8 | 0,153287028 | 1,081920004 | 1,367336946 | 0 |
| 9 | 0,043906853 | 1,045191167 | 1,32412867 | 0 |
| 10 | 0,093573226 | 1,074726965 | 1,359582549 | 0 |
| ... | ... | ... | ... | ... |
| 25 | 0,021093222 | 1,054890669 | 1,335410693 | 0 |

Setelah perhitungan menggunakan *Euclidean Distance* dilakukan, data dikelompokkan berdasarkan jarak minimum ke centroid terdekat.

cluster lama (0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0)

cluster baru (0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0)

Terjadi perubahan pengelompokkan data maka akan dilanjutkan ke perhitungan iterasi berikutnya.

Sebelum melanjutkan ke iterasi ke-3, peneliti perlu menentukan centroid baru dengan menghitung nilai rata-rata setiap atribut pada data yang termasuk dalam cluster 0, cluster 1 dan cluster 2.

Setelah dilakukan perhitungan untuk mencari centroid baru di dapat hasil sebagai berikut.

$$C0 = (0,196034272; 0,227481019; 0; 0)$$

$$C1 = (0,4521275; 0,4340275; 1; 0)$$

$$C2 = (0,417794545; 0,598483636; 0,363636364; 1)$$

Iterasi 3

Tabel 8. Hasil Iterasi 3

| Data | Jarak Centroid Itrasi 3 | | | Cluster |
|------|-------------------------|-------------|-------------|---------|
| | C0 | C1 | C2 | |
| 1 | 0,063043024 | 1,035362467 | 1,310572238 | 0 |
| 2 | 0,134971313 | 1,09979765 | 1,387468744 | 0 |
| 3 | 0,192968665 | 1,023652011 | 1,286592524 | 0 |
| 4 | 0,094978019 | 1,027240568 | 1,300465239 | 0 |
| 5 | 0,026354721 | 1,060348949 | 1,341523775 | 0 |
| 6 | 0,066067186 | 1,072826878 | 1,357044182 | 0 |
| 7 | 0,118887408 | 1,065854227 | 1,34273726 | 0 |
| 8 | 0,149076999 | 1,081920004 | 1,367336946 | 0 |
| 9 | 0,038368134 | 1,045191167 | 1,32412867 | 0 |
| 10 | 0,088707492 | 1,074726965 | 1,359582549 | 0 |
| ... | ... | ... | ... | ... |
| 25 | 0,006951887 | 1,054890669 | 1,335410693 | 0 |

Setelah perhitungan menggunakan *Euclidean Distance* dilakukan, data dikelompokkan berdasarkan jarak minimum ke centroid terdekat.

cluster lama (0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0)

cluster baru (0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0)

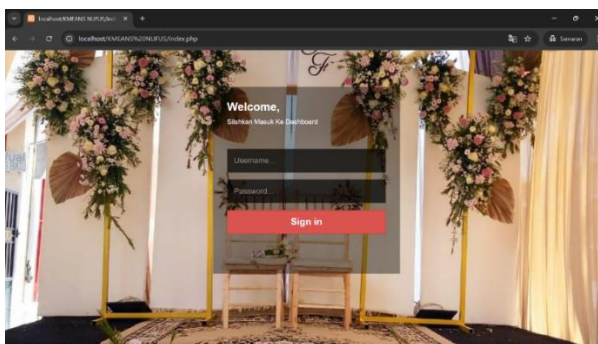
Tidak terjadi perubahan pengelompokkan data, maka proses iterasi dianggap selesai. Pada tahap ini, algoritma K-Means telah mencapai konvergensi, di mana posisi centroid dan pengelompokkan data telah stabil. Hasil akhir ini menunjukkan bahwa kluster telah terbentuk dengan optimal, dan data telah dikelompokkan sesuai dengan karakteristik yang diidentifikasi dalam analisis.

Hasil akhir dari perhitungan iterasi menunjukkan bahwa jumlah data dalam masing-masing kluster adalah sebagai berikut: C0 memiliki 206 data, C1 memiliki 8 data, dan C2 memiliki 7 data. Yang dapat dilihat pada tabel berikut.

Tabel 9. Jumlah Akhir Cluster

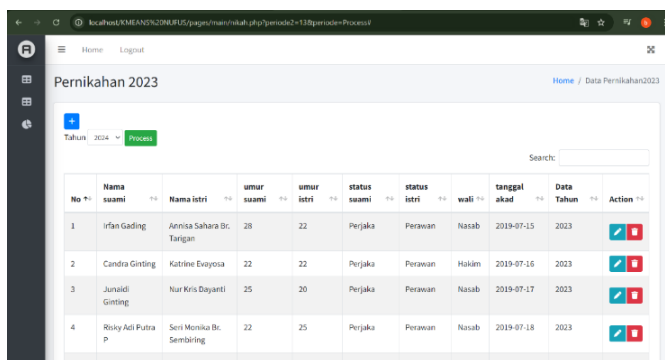
| Jumlah Cluster | | |
|----------------|----|----|
| C0 | C1 | C2 |
| 206 | 8 | 11 |

3.6 Penerapan Pengelompokan Data di Web



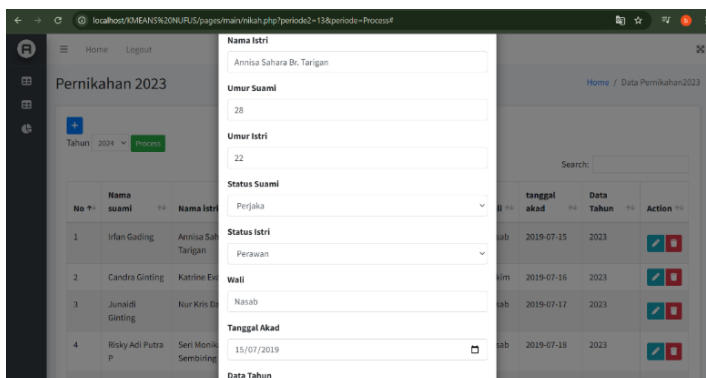
Gambar 3. Tampilan Login

Gambar 3 adalah tampilan *login*. Halaman *login* ini memiliki dua kolom *input* di bagian tengah: kolom pertama untuk memasukkan '*Username*', dan kolom kedua untuk memasukkan '*Password*'. Setelah memasukkan informasi *login*, pengguna dapat menekan tombol '*Sign in*' untuk mengakses aplikasi. Halaman ini didesain dengan antarmuka yang bersih dan mudah digunakan, dengan latar belakang yang sederhana dan profesional.



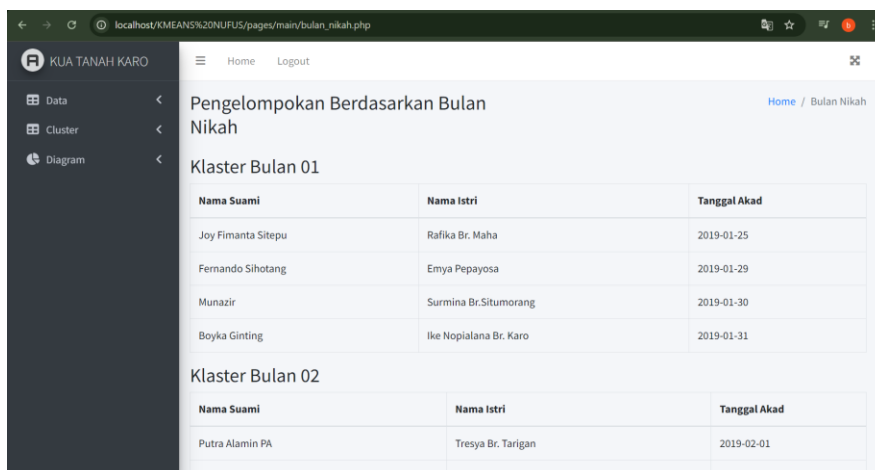
Gambar 4. Tampilan Data

Gambar 4 adalah tampilan data. Halaman ini menunjukkan tampilan data yang telah berhasil di *input* dan di simpan dalam sistem. Data ditampilkan dalam format tabel, dengan setiap entri mencakup informasi yang telah dimasukkan sebelumnya, seperti Nama Suami dan Istri, umur, status, wali, tanggal akad, dan data tahun. Tampilan ini juga dilengkapi dengan opsi untuk mengedit atau menghapus data, yang memungkinkan pengguna untuk memodifikasi informasi jika diperlukan. Dengan desain yang jelas dan terstruktur, pengguna dapat dengan mudah melihat dan memverifikasi data yang telah di *input*.



Gambar 5. Tampilan Formulir

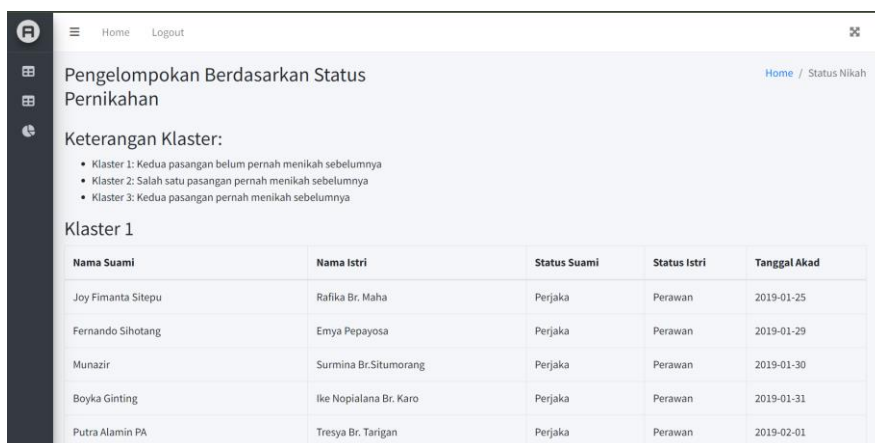
Gambar 5 ini menunjukkan tampilan formulir untuk menambahkan data baru ke dalam sistem. Formulir ini terdiri dari beberapa kolom input, di mana pengguna dapat memasukkan informasi yang diperlukan, seperti tahun, jumlah pernikahan, atau atribut lainnya yang relevan.



| Klaster Bulan 01 | | |
|--------------------|------------------------|--------------|
| Nama Suami | Nama Istri | Tanggal Akad |
| Joy Fimanta Sitepu | Rafika Br. Maha | 2019-01-25 |
| Fernando Sihotang | Emya Pepayosa | 2019-01-29 |
| Munazir | Surmina Br.Situmorang | 2019-01-30 |
| Boyka Ginting | Ike Nopialana Br. Karo | 2019-01-31 |
| Klaster Bulan 02 | | |
| Nama Suami | Nama Istri | Tanggal Akad |
| Putra Alamin PA | Tresya Br. Tarigan | 2019-02-01 |

Gambar 6. Tampilan Cluster Bulan Pernikahan

Gambar 6 adalah tampilan cluster bulan pernikahan. Halaman ini menunjukkan tampilan hasil *clustering* yang dikelompokkan berdasarkan bulan. Data yang ditampilkan terbagi dalam beberapa *cluster*, dengan setiap cluster mewakili kelompok informasi yang relevan yang dianalisis berdasarkan bulan tertentu.



Keterangan Klaster:

- Klaster 1: Kedua pasangan belum pernah menikah sebelumnya
- Klaster 2: Salah satu pasangan pernah menikah sebelumnya
- Klaster 3: Kedua pasangan pernah menikah sebelumnya

| Klaster 1 | | | | |
|--------------------|------------------------|--------------|--------------|--------------|
| Nama Suami | Nama Istri | Status Suami | Status Istri | Tanggal Akad |
| Joy Fimanta Sitepu | Rafika Br. Maha | Perjaka | Perawan | 2019-01-25 |
| Fernando Sihotang | Emya Pepayosa | Perjaka | Perawan | 2019-01-29 |
| Munazir | Surmina Br.Situmorang | Perjaka | Perawan | 2019-01-30 |
| Boyka Ginting | Ike Nopialana Br. Karo | Perjaka | Perawan | 2019-01-31 |
| Putra Alamin PA | Tresya Br. Tarigan | Perjaka | Perawan | 2019-02-01 |

Gambar 7. Tampilan Cluster Status Pernikahan

Gambar 7 adalah tampilan cluster status pernikahan. Pada halaman ini menunjukkan tampilan visualisasi cluster yang dibagi berdasarkan status pernikahan.

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa metode elbow menghasilkan tiga cluster yang terbentuk. Pembagian cluster tersebut adalah sebagai berikut: Cluster 0 (C0) berisi 206 data, Cluster 1 (C1) terdiri dari 8 data, dan Cluster 2 (C2) mencakup 11 data. Hasil ini menunjukkan bahwa sebagian besar data terkelompok dalam Cluster 0, sementara Cluster 1 dan Cluster 2 berisi jumlah data yang lebih kecil, mencerminkan distribusi yang tidak merata di antara ketiga cluster tersebut. Setelah mendapatkan 3 cluster, kita bisa menyimpulkan beberapa hal berdasarkan interpretasi cluster tersebut: **Cluster 1:** Mungkin berisi "Calon Suami" dan "Calon Istri" yang memiliki umur dan status yang relatif sama. **Cluster 2:** Mungkin berisi pasangan dengan perbedaan umur yang cukup signifikan atau dengan status yang sangat berbeda. **Cluster 3:** Bisa berisi kelompok dengan status atau umur yang lebih beragam, atau mungkin kelompok dengan perbedaan ekstrem dalam satu atau dua atribut.

REFERENCES

[1] Ahmad Agung Setya Budi, "Kawin Paksa Dalam Perspektif Hukum Islam Dan Konteks Kajian Hak Asasi Manusia," *Jurnal Dunia Ilmu Hukum (JURDIKUM)*, vol. 1, no. 2, pp. 44–49, Sep. 2023, doi: 10.59435/juridikum.v1i2.168.

- [2] B. Ulum and A. Muzawwir, “Analisis Pertimbangan Hakim Lama Pacaran Sebagai Alasan Mendesak Mengabulkan Permohonan Dispensasi Nikah Dini Study Putusan Nomor: 354/Pdt. P/2022/Pa Bangkalan,” *Al-Ibrah: Jurnal Pendidikan dan Keilmuan Islam*, vol. 8, no. 2, pp. 92–111, 2023.
- [3] A. A. Musyafah, J. Sudarto, and J. Tengah, “Perkawinan Dalam Perspektif Filosofis Hukum Islam,” *Credito*, no. 2, pp. 111–122, 2020, doi: 10.14710/nts.v12i1.28897.
- [4] Kevin Marpaung, “Penerapan Sistem Informasi Manajemen Nikah (Simkah) Berbasis Web Di Kantor Urusan Agama Kecamatan Kedaton Kota Bandar Lampung,” Universitas Lampung, Bandar Lampung, 2022.
- [5] H. R. U. Sembiring, *Membangun Pribadi Prima Dalam Pelayanan Publik*. Media Nusa Creative (MNC Publishing), 2021.
- [6] S. Syam *et al.*, *Data Mining: Teori dan Penerapannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2024.
- [7] M. T. Jatipaningrum, S. E. Azhari, and K. Suryowati, “Pengelompokan Kabupaten Dan Kota Di Provinsi Jawa Timur Berdasarkan Tingkat Kesejahteraan Dengan Metode K-Means Dan Density-Based Spatial Clustering Of Applications With Noise,” *Jurnal Derivat*, vol. 9, no. 1, 2022.
- [8] Ramadhana, Islamiyah, and A. P. A. Masa, “Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara,” *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 2, no. 1, pp. 35–42, Jun. 2023, doi: 10.30872/atasi.v2i1.595.
- [9] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. I. P. Adab, *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab, 2023.
- [10] H. Astuti, “Penerapan data mining menggunakan metode k-means clustering untuk mengelompokkan data pelanggan (studi kasus: pt. pinus merah abadi),” *Jurnal Web Informatika Teknologi*, vol. 6, no. 1, pp. 1–8, 2021.
- [11] A. Nugraha, O. Nurdiawan, and G. Dwilestari, “Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport,” *Jurnal Mahasiswa Teknik Informatika*, vol. 6, no. 2, pp. 849–855, 2022.
- [12] H. Priyatman, F. Sajid, and D. Haldivany, “JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa,” vol. 5, no. 1, pp. 62–66, 2019.
- [13] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [14] L. Petra Refialy, H. Maitimu, and M. Soyano Pesulima, “Perbaikan Kinerja Clustering K-Means pada Data Ekonomi Nelayan dengan Perhitungan Sum of Square Error (SSE) dan Optimasi nilai K cluster,” *Techno.COM*, vol. 20, no. 2, pp. 321–329, 2021.
- [15] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, p. 100, 2021.
- [16] N. U. R. H. Tambunan, “Analisis Data Mining Clustering Tentang Tingkat Kesejahteraan Di Desa Karang Anyar Menggunakan Metode K-Means,” 2024, *Universitas Labuhanbatu*.
- [17] I. Y. Beti and H. Juliansa, “Penerapan Normalisasi Data Metode Decimal Scaling Dan Metode K-Means Dalam Mengelompokkan Kasus Demam Berdarah,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 2928–2936, 2024.
- [18] R. F. Putra *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023.
- [19] F. Indriyani and E. Irfiani, “Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means (Clustering Sales Data at Outdoor Equipment Stores Using K- Means Method),” *JUITA: Jurnal Informatika*, vol. 7, no. 2, pp. 109–113, 2019.
- [20] R. Kinanti, “Penerapan Metode Clustering K-Means untuk Menentukan Prioritas Penerima Bantuan Program Beras untuk Rakyat Miskin (Raskin) Studi Kasus: Kecamatan Siulak,” *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 4, no. 2, 2024, doi: 10.33998/jakakom.v4i2.