KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN METODE *K-MEANS*

SKRIPSI

Untuk Memenuhi Persyaratan
Memperoleh Gelar Sarjana

Oleh:

Malinda Mahanani
20.51.0008

PROGRAM STUDI S1 – SISTEM INFORMASI

KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
PPKIA PRADNYA PARAMITA
MALANG
2024

LEMBAR PERSETUJUAN

KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN METODE *K-MEANS*

Oleh:
Malinda Mahanani
20.51.0008

Telah disetujui oleh pembimbing untuk diujikan
Pada tanggal 6/5/2024

Menyetujui,
Komisi Pembimbing

Pembimbing I	Pembimbing II
Dwi Safiroh Utsalina, S.Kom., M.M.S.I	Dr. Tb. Mohammad Akhriza, S.Si, M.M.S.I
NIK 00.12.01.016	NIK 00.12.01.002

Mengetahui,
Ketua Program Studi S-1 Sistem Informasi

Dwi Safiroh Utsalina, S.Kom., M.M.S.I
NIK 00.12.01.016

LEMBAR PENGESAHAN

KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN METODE *K - MEANS*

Skripsi oleh Malinda Mahanani
Telah dipertahankan di depan dewan penguji
Pada tanggal 5 Mei 2024

Dewan Penguji,
Ketua

Dr. Weda Adistianaya Dewa, S.Kom.,MMSI
NIK 14.02.25.003

Anggota I

Eka Yuniar, S.Kom., MMSI
NIK 16.09.19.001

Anggota II

Dr.Tb. Mohammad Akhriza,S.Si.,MMSI
NIK 00.12.01.002

Mengetahui,
Ketua Program Studi
S-1 Sistem Informasi

Dr. Dwi Saftiah Utsalina, S.Kom., M.M.S.I NIK 00.12.01.016

Mengesahkan,
Ketua STMIK PPKIA
Pradnya Paramita

Dr. Tb. Mohammad Akhriza, S.Si, M.M.S.I
NIK 00.12.01.002

PERNYATAAN
ORISINALITAS SKRIPSI

Saya menyatakan bahwa tugas akhir yang telah saya buat dengan judul: "Klasterisasi Data Perceraian Di Wilayah Kabupaten Lamongan Dengan Menggunakan metode K-Means", adalah asli (orisinil) atau tidak plagiat (menjiplak) dan belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun, kecuali yang secara tertulis dikutip dalam naskah ini disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang saya peroleh (SARJANA) dibatalkan,serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No 20 Tahun 2003, pasal 25 ayat 2 dan pasal 70).

Malang, 4 Mei 2024

Malinda Mahanani

NIM 20.51.0008

Motto Hidup

“Semua Berasal dari Allah dan Akan Kembali Kepada-Nya”

“Jika Kau Tidak Mampu Terbang, Maka berlarilah. Hari ini Kita Akan Bertahan. Jika kau Tak Mampu Berlari, Maka Berjalanlah.”

ABSTRAK

Mahanani, Malinda Mahanani. 2024. *Klusterisasi Data Perceraian di Wilayah Kabupaten Lamongan Dengan Menggunakan Metode K-means*. Program Studi Sistem Informasi STMIK PPKIA Pradnya Paramita. Pembimbing: (1) Dr. Dwi Safiroh Utsalina, S.Kom., MMSI, (II) Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D.

Kata Kunci: *Data Mining, Perceraian,, Cluster, K-Means.*

Perceraian adalah terminasi dari ikatan pernikahan yang terjadi ketika pasangan memutuskan untuk mengakhiri kehidupan pernikahan mereka. Hal ini disebabkan oleh berbagai faktor, termasuk ketidakmampuan suami dan istri untuk menjalankan peran masing-masing dengan baik. Kasus perceraian meningkat dari tahun ke tahun, termasuk di daerah Lamongan. Penelitian ini dilakukan untuk menganalisis data perceraian yang terjadi di daerah Lamongan dengan menggunakan metode klasterisasi K-means. Analisis ini didasarkan pada atribut jenis cerai, umur penggugat, umur tergugat, faktor penyebab, dan KUA tempat nikah/kecamatan. untuk mengidentifikasi daerah yang memiliki tingkat perceraian tertinggi di Kabupaten Lamongan pada tahun 2020 - 2022. Bertujuan untuk Melakukan analisis data perceraian di pengadilan agama lamongan dengan melakukan klasterisasi K-means guna mengelompokan daerah yang memiliki kasus perceraian tertinggi berdasarkan karakteristik yang ditentukan.
Berdasarkan hasil nilai silhoutte score yang diperoleh data tahun 2020 yang bernilai dengan rentan 0,39 – 0,53 tahun 2021 bernilai dengan rentan 0,50 – 0,61, dan pada tahun 2022 bernilai dengan rentan 0,37 – 0,53. Adapun analisis posisi siku pada grafik elbow menunjukkan penurunan signifikan dalam nilai inertia seiring dengan peningkatan jumlah cluster. Inertia mengukur seberapa jauh titik data tersebar dalam sebuah cluster. Inertia cenderung menurun saat jumlah cluster bertambah, tetapi penurunan tersebut melambat setelah mencapai jumlah cluster optimal. Titik di mana penurunan inertia menjadi kurang signifikan menandakan jumlah cluster optimal. Hasil analisis menunjukkan bahwa terbagi menjadi empat klaster, data perceraian di kecamatan Babat, Paciran, dan Brondong menunjukkan bahwa kasus perceraian tertinggi disebabkan oleh faktor ekonomi, perselisihan, dan pertengkaran terus-menerus. Pada tahun 2022, terjadi peningkatan kasus perceraian di mana perselisihan dan pertengkaran menjadi penyebab utama.

KATA PENGANTAR

Puji syukur kami panjatkan kehadirat Tuhan Yang Maha Esa atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi dengan judul "KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN ALGORITMA K-MEANS dengan baik sebagai syarat untuk memperoleh gelar Sarjana Komputer..

Kami menyadari tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan skripsi ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Tuhan Yang Maha Esa yang telah memberikan petunjuk dan hidayah dalam pembuatan skripsi dan penyusunan laporan sehingga dapat berjalan dengan baik dari awal hingga akhir.
2. Kedua orang tua saya yang telah memberikan doa dan dukungannya.
3. Bapak Dr. Tb. Mohammad Akhriza, S.Si, M.M.S.I, selaku Ketua STMIK PPKIA Pradnya Paramita
4. Ibu Dwi Safiroh Utsalina, S.Kom., M.M.S.I., selaku Dosen Pembimbing 1 Sistem Informasi dan ketua program Studi S-1 Sistem informasi
5. Bapak Dr. Tb. Mohammad Akhriza, S.Si, M.M.S.I, selaku dosen pembimbing 2 Sistem Informasi
6. Bapak hakim dan staf pengadilan agama lamongan yang telah memberikan dukungan dalam proses pengambilan data.
7. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Skripsi dari awal hingga akhir yang tidak dapat kami sebutkan satu persatu.

Penulis mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan skripsi ini. Semoga skripsi ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 4 Mei 2024

Malinda Mahanani

NIM 20.51.0008

DAFTAR ISI

JUDUL SKRIPSI i

LEMBAR PERSETUJUAN ii

LEMBAR PENGESAHAN iii

PERNYATAAN ORISINALITAS SKRIPSI iv

ABSTRAK vi

KATA PENGANTAR vii

DAFTAR ISI ix

DAFTAR TABEL xi

DAFTAR GAMBAR xii

DAFTAR LAMPIRAN xv

BAB I PENDAHULUAN 1

1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 4
1.5 Kontribusi Penelitian 5

BAB II TINJAUAN PUSTAKA 6

2.1 Data Mining 6
2.2 Metode Clastering 9
2.3 Algoritma K-means dan Metode Elbow 10
2.4 Penerapan K-Means Pada Kasus Perceraian 13
2.5 Evaluasi Clustering 15

BAB III METODE PENELITIAN 16

3.1 Analisis Permasalahan 16
3.2 Solusi yang Diusulkan 17

3.2.1 Seleksi Fitur 18
3.2.2 Preprocessing 19
3.2.3 Menentukan Jumlah Cluster Optimal 21
3.2.4 Evaluasi dan Hasil Analisis Cluster 22
3.2 Rancangan Eksperimen 22
3.3.1 Peralatan 22
3.3.2 Kebutuhan Data 23
BAB IV PENGUJIAN DAN HASIL 24
4.1 Penerapan Metode 24
4.2.1 Seleksi Fitur 24
4.2.2 Pre-Processing 25
4.2.3 Cleaning Data 25
4.2.4 Transformasi Data 26
4.2.5 Normalisasi Data 29
4.2.6 Menentukan Jumlah Cluster Optimal 35
4.2.7 Menjalankan Kmeans Clustering 41
4.2 Evaluasi dan Hasil Analisis Cluster 44
4.2.1 Evaluasi Cluster 44
4.2.1 Hasil Analisis Cluster 44
BAB V KESIMPULAN DAN SARAN 62
5.1 Kesimpulan 62
5.2 Saran 63
DAFTAR PUSTAKA

DAFTAR TABEL

No Halaman

Tabel 4. 1 Sample Data Mentah Penceraian 2020 24

Tabel 4. 2 Sample Data Mentah Penceraian 2021 24

Tabel 4. 3 Sample Data Mentah Penceraian 2022 25

Tabel 4. 4 Hasil Clustering Data 2020 44

Tabel 4. 5 Hasil Clustering Data 2021 46

Tabel 4. 6 Hasil Clustering Data 2022 48

DAFTAR GAMBAR

No Halaman

Gambar 2. 1 Proses Data Maining 7

Gambar 2. 2 Flowchart Algoritma K-Means 11

Gambar 3. 1 Flowchart Penelitian..... 18

Gambar 4. 1 Source Code Cleansing 25

Gambar 4. 2 Hasil Cleansing 25

Gambar 4. 3 Source Code Import Data 2020..... 26

Gambar 4. 4 Souce Code Import Data 2021 26

Gambar 4. 5 Source Code Import Data 2022..... 26

Gambar 4. 6 Hasil Import Data Set 2020..... 27

Gambar 4. 7 Hasil Import Data Set 2021..... 27

Gambar 4. 8 Hasil Import Data Set 2022..... 27

Gambar 4. 9 Source Code Transformasi Data 28

Gambar 4. 10 Hasil Transformasi Data 2020..... 28

Gambar 4. 11 Hasil Transformasi Data 2021..... 29

Gambar 4. 12 Hasil Transformasi Data 2022..... 29

Gambar 4. 13 Hasil Normalisasi Data Tahun 2020 32

Gambar 4. 14 Hasil Normalisasi Data Tahun 2021 33

Gambar 4. 15 Hasil Normalisasi Data Tahun 2022 33

Gambar 4. 16 Source Code Normalisasi Data 34

Gambar 4. 17 Hasil Normalisasi Data 2020 34

Gambar 4. 18 Hasil Normalisasi Data 2021 35

Gambar 4. 19 Hasil Normalisasi Data 2022 35

Gambar 4. 20 Hasil Source Code Jumlah Optimal 36
Gambar 4. 21 Hasil dari Silhoutte_score Data 2020..... 37
Gambar 4. 22 Hasil dari Silhoutte_score Data 2021..... 37
Gambar 4. 23 Hasil dari Silhoutte_score Data 2022..... 37
Gambar 4. 24 Hasil Visualisasi Data Jumlah Optimal 2020..... 38
Gambar 4. 25 Hasil Visualisasi Data Jumlah Optimal 2020..... 38
Gambar 4. 26 Hasil Visualisasi Data Jumlah Optimal 2021..... 39
Gambar 4. 27 Hasil Visualisasi Data Jumlah Optimal 2021..... 39
Gambar 4. 28 Hasil Visualisasi Data Jumlah Optimal 2022..... 40
Gambar 4. 29 Hasil Visualisasi Data Jumlah Optimal 2022..... 40
Gambar 4. 30 Sourcing Code Clustering 42
Gambar 4. 31 Hasil Clustering Data Tahun 2020..... 42
Gambar 4. 32 Hasil Clustering Data Tahun 2021..... 43
Gambar 4. 33 Hasil Clustering Data Tahun 2022..... 43
Gambar 4. 34 hasil chart claster 1..... 50
Gambar 4. 35 hasil chart claster 2..... 51
Gambar 4. 36 hasil chart claster 3..... 52
Gambar 4. 37 hasil chart claster 4..... 53
Gambar 4. 38 hasil chart claster 1..... 54
Gambar 4. 39 hasil chart claster 2..... 55
Gambar 4. 40 hasil chart claster 3..... 56
Gambar 4. 41 hasil chart claster 4..... 57
Gambar 4. 42 hasil chart claster 1..... 58
Gambar 4. 43 hasil chart claster 2..... 59

Gambar 4. 44 hasil chart claster 3 60

Gambar 4. 45 hasil chart claster 4 61

DAFTAR LAMPIRAN

1. Riwayat Hidup
2. Berita Acara Bimbingan

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perceraian merupakan putusnya hubungan antara suami dan istri dalam ikatan pernikahan yang sah. Putusnya pernikahan suami dan istri terjadi atas kesepakatan kedua nya apabila hubungan mereka tidak lagi memungkinkan tercapainya tujuan perkawinan. Oleh karena itu, perceraian dapat dilakukan apabila dengan alasan yang kuat dengan hukuman perkawinan yang berlaku di Indonesia di tuangkan di dalam Undang – Undang 1 Tahun 1974 dan Peraturan Pemerintah Nomor 9 Tahun 1975. (nur & dkk, 2022) .Menurut (pawestri & hillary sekar, 2023) terdapat beberapa efek perceraian diantaranya menimbulkan stress sampai dengan depresi, stress yang dialami bahkan dapat meningkatkan resiko penyakit jantung dan stroke. Selain dampak pada pasangan yang berpisah, perceraian juga meninggalkan luka psikologis yang disebut trauma bagi anak-anak dari pasangan tersebut (amelia & ema, 2023).

Tingginya tingkat perceraian di Indonesia mencapai angka 516,000 pada tahun 2022, bahkan diketahui bahwa angka tersebut merupakan peningkatan 15% dari tahun sebelumnya (hidayah & Nur Fitri, 2023). Sehingga dengan terus meningkatnya kasus perceraia, berakibat negatif pada usaha pemerintah dalam mempersiapkan generasi penerus bangsa yang siap manghadapi segala macam tantangan hidup di masa mendatang.

Kabupaten Lamongan merupakan salah satu wilayah yang memiliki 29 kecamtan dengan tingkat kasus perceraian relatif tinggi yaitu pada tahun

2685 (CATAHU 2020), 2683 (CATAHU 2021), 2608 (CATAHU,2022). Sebagai lembaga peradilan, Pengadilan Agama Lamongan memiliki fungsi untuk melayani permintaan dari lembaga lain. Permintaan tersebut termasuk permintaan data yang telah terkumpul di Pengadilan Agama Lamongan. Salah satu jenis data yang berasal dari Pengadilan Agama Lamongan yaitu data kasus perkara perceraian. Berdasarkan hasil wawancara pada tanggal 26 Juli 2023 di pengadilan agama setempat (Lamongan), diketahui bahwa kasus perceraian yang diputus di pengadilan Agama setempat hanya digunakan untuk pembuatan laporan. Dalam rangka mengoptimalkan pengelolaan data yang ada kekhawatiran terus meningkatnya kasus perceraian di Indonesia.

Dalam rangka mengoptimalkan pengelolaan data yang ada, misalnya untuk mengetahui tren perceraian, mengelompokkan atribut yang disesuaikan dengan kondisi dan situasi pada obyek penelitian (Haris , Sarjon , & Sumijan, 2020). Oleh karena itu diperlukan sebuah teknik atau metode untuk mengelompokkannya. Klastering merupakan pendekatan yang sesuai untuk mengelompokkan data yang tidak memiliki label, dari beragam algoritma Klastering yang ada, diketahui bahwa algortima K-Means lebih sesuai digunakan untuk permasalahan di penelitian ini, karena menurut (insanul, ulya, & mustakin, 2019), bahwa algoritma K-Means lebih cepat dibanding algoritma K-Medoids dalam memproses data. Penelitian mengenai pengelompokkan perceraian pernah dilakukan oleh Uci Dwi Rahayu, Nuke Chusna, Moch Fachri pada tahun 2022, (uci, Nuke , & Moch , 2022) Penelitian serupa pernah dilakukan oleh N Nurhayati1, Fahrija Azzahra, Sri Ramadani, Sinta Dwi Hastuti, Eka Irawan pada tahun 2020, (nur & dkk,

2022) Atribut atau fitur yang dikelompokkan yaitu pada perceraian pada pengadilan Bekasi. Kasus penyebab perceraian berdasarkan provinsi. Meskipun sudah banyak penelitian yang telah dilakukan mengenai indikator tingkat perceraian, namun dalam penelusuran melalui *Google Scholar* dalam publikasi berbahasa Indonesia masih jarang penelitian menggunakan klastering dengan objek seputar kasus perceraian. Penelitian ini menggunakan klastering dengan beberapa atribut yaitu Jenis cerai, Umur-p, Umur-t, Faktor Perceraian, KUA Tempat Nikah.

Berdasarkan permasalahan yang telah disampaikan, maka penelitian ini mengelompokkan data perceraian tahun 2022 yang paling tinggi di di kecamatan-kecamatan yang berada dilingkungan kabupaten Lamongan dengan judul " KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN ALGORITMA K-MEANS".

1.2 Rumusan Masalah

Melihat dari latar belakang yang sudah dijabarkan rumusan masalahnya adalah.

1. Bagaimana cara pengadilan mengetahui angka perceraian tertinggi di berbagai kecamatan kabupaten lamongan berdasarkan karakteristik yang di tentukan dengan menggunakan metode k-means
2. Bagaimana kualitas klastering yang dihasilkan apabila menggunakan metode Silhoutte Coefficient.
3. Bagaimana kualitas menentukan clastering data perceraian pada tahun

2020 - 2022 berdasarkan atribut jenis cerai, umur penggugat, umur tergugat, faktor penyebab, dan kua tempat nikah

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah :

1. Melakukan analisis data perceraian di pengadilan agama lamongan dengan melakukan klasterisasi K-means guna mengelompokan daerah yang memiliki kasus perceraian tertinggi berdasarkan karakteristik yang ditentukan.
2. Mengetahui kualitas hasil klastering dengan menggunakan Silhoutte Coefficient.
3. Mengetahui clastering data perceraian pada tahun 2020 - 2022 berdasarkan atribut jenis cerai, umur penggugat, umur tergugat, faktor penyebab, dan kua tempat nikah.

1.4 Batasan Masalah

Batasan masalah pada penelitian data perceraian pengadilan agama lamongan dengan mengunakan metode K-Means adalah :

1. Data perceraian yang akan digunakan dalam penelitian ini terdiri dari terbitnya akte cerai rentang waktu tahun 2020 - 2022.
2. Data perceraian yang akan digunakan meliputi variabel-variabel tertentu, seperti jenis cerai, umur penggugat, umur tergugat, faktor perceraian, dan KUA tepat menikah Penelitian ini akan menggunakan Metode K-Means, untuk data perceraian menjadi kelompok-kelompok tertentu.

1.5 Kontribusi Penelitian

Kontribusi klasterisasi data perceraian untuk Pengadilan Agama Lamongan dengan menggunakan Metode K-Means

1. Dengan menerapkan Metode K-Means penelitian ini dapat membantu mengelompokkan data perceraian menjadi kelompok-kelompok dengan nilai *k* terbaik dari perhitungan metode Elbow serta perhitungan kualitas clustering yang dihasilkan menggunakan metode Silhoutte Coefficient.

2. Hasil penelitian ini dapat digunakan untuk memberdayakan masyarakat setempat dengan memberikan informasi yang lebih baik tentang kondisi perceraian di wilayah Lamongan sehingga masyarakat dapat lebih terinformasi dan sadar isu-isu tersebut.

BAB II
TINJAUAN PUSTAKA

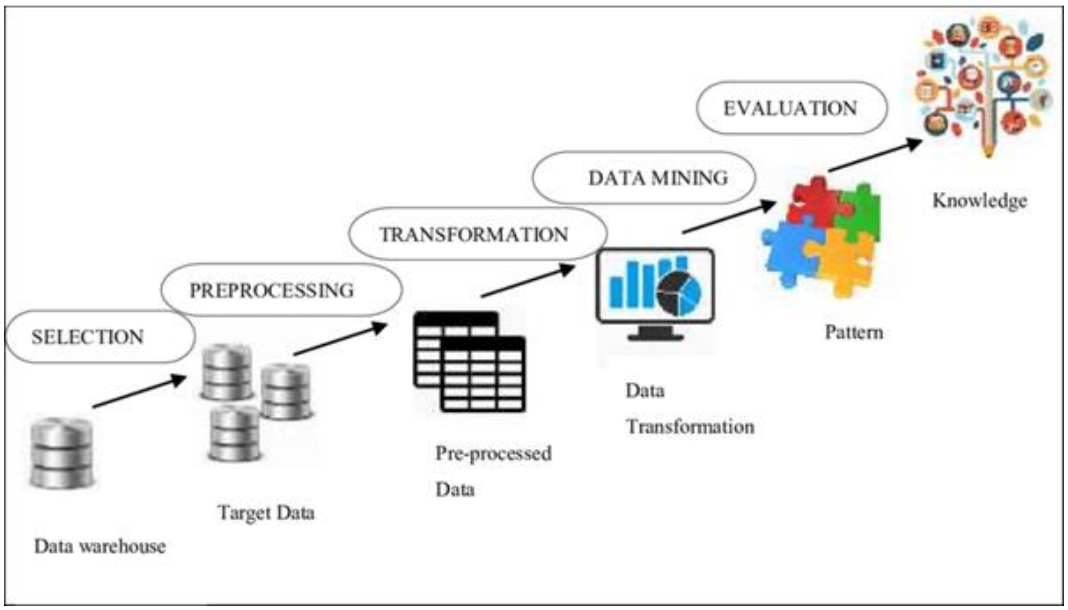
2.1 Data Mining

Data mining adalah proses yang terkait dengan ekstraksi pengetahuan dari dalam pengumpulan data (buulolo & efori, 2020). Data mining dapat dijelaskan sebagai proses pengumpulan dan analisis data yang bertujuan untuk mengekstrak informasi berharga dari dalam dataset. Proses pengumpulan dan ekstraksi informasi ini dapat diperlakukan dengan menggunakan perangkat lunak yang menggunakan teknik statistika, matematika, dan teknologi kecerdasan buatan (AI), dapat melakukan deskripsi pada kumpulan data target atau melakukan prediksi hasil dengan menggunakan algoritma machine learning. Data mining seringkali disebut juga sebagai Knowledge Discovery in Database (Setiawan & Rony, 2021). Manfaat Data mining untuk memastikan pengumpulan dan analisis data yang dapat diandalkan. Ini sering melibatkan proses yang terstruktur dan formal dalam mengidentifikasi masalah, mengumpulkan data yang relevan, dan berusaha merumuskan solusi (jacob & jessica, 2023).

Ada beberapa tujuan Data Mining yaitu *explanatory* yang digunakan untuk menjelaskan suatu kondisi, *confirmatory* yang digunakan untuk mengklarifikasi sebuah pernyataan adanya hipotesa, dan exploratory yang digunakan untuk menemukan pola baru yang sebelumnya tidak terdeteksi (Nugroho & Andy, 2020).

Menurut sumber yang dikutip (binus, 2021) dalam ranah Data Mining, terdapat sejumlah proses yang perlu dilalui dan dijalankan dengan

seksama untuk meraih pemahaman yang lebih mendalam dan berarti dari suatu dataset yang cenderung besar dan kompleks, seperti pada Gambar 2.1 .



Gambar 2. 1 Proses Data Maining

Berikut ini penjelasan dari tahapan proses data mining pada gambar 2 1 :

1. Data Cleansing adalah proses di mana data diolah, dan kemudian dipilih data yang dianggap dapat digunakan dengan cara membuang atau memperbaiki data yang tidak akurat, tidak lengkap, atau tidak relevan. Seperti contoh pada kasus data perceraian, data yang tidak di butuhkan untuk proses pengelompokan daerah yang tinggi Tingkat perceraian maka data yang tidak berkaitan, seperti misal data latar belakang Pendidikan tidak relevan atau tidak diperluka lebih baik dihapus dari daftar, guna meningkatkan kualitas dan ketepatan data, sedangkan data yang berkesesuaian, seperti jenis perceraian, usia p, usia t, faktor penyebab, dan tempat KUA/Kecamatan bisa tetap dimunculkan.
2. Data Integration merupakan proses menggabungkan data yang dianggap berulang menjadi satu kesatuan dengan tujuan menciptakan tampilan yang komprehensif dan terpadu dari berbagai sumber data. Dalam proses ini, data dari

berbagai sumber yang mungkin memiliki format, struktur, atau sistem yang berbeda diintegrasikan sehingga dapat diakses dan dianalisis secara efisien.

3. Selection merupakan proses seleksi atau pemilihan data yang dianggap relevan untuk analisis, di mana hanya data yang memenuhi kriteria tertentu yang dipilih untuk dimasukkan ke dalam dataset akhir.
4. Data Transformation merupakan proses mengubah data yang terpilih ke dalam bentuk prosedur penambangan (mining procedure) dengan maksud untuk mempersiapkan dataset sehingga dapat diolah dan dianalisis lebih lanjut. Dalam tahap ini, data yang telah dipilih melalui proses seleksi dapat mengalami transformasi berupa normalisasi, pengkodean ulang, atau penggabungan kolom untuk menciptakan struktur data yang lebih sesuai dengan kebutuhan analisis atau model penambangan data yang diterapkan.
5. Data Mining merupakan proses di mana beragam teknik digunakan untuk mengekstrak pola-pola potensial dari data dengan tujuan menghasilkan informasi yang berguna. Melalui penerapan algoritma dan model analisis data, data mining dapat mengidentifikasi hubungan, tren, atau pola yang mungkin tidak terlihat secara langsung..
6. Pattern Evolution merupakan proses di mana pola-pola yang telah diidentifikasi dievaluasi berdasarkan pengukuran (measure) yang diberikan. Dalam tahap ini, pola-pola yang ditemukan melalui analisis data atau data mining dianalisis lebih lanjut untuk memahami sejauh mana relevansi dan signifikansinya. Pengukuran yang diberikan dapat mencakup aspek seperti tingkat kepercayaan (confidence), tingkat dukungan (support), atau metrik lain yang sesuai dengan tujuan analisis.

7. Knowledge Presentation merupakan tahap terakhir dalam proses KDD (Knowledge Discovery in Databases), di mana data yang telah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan memungkinkan mereka untuk mengambil tindakan berdasarkan analisis tersebut. Dalam tahap ini, wawasan dan temuan yang dihasilkan dari analisis data disajikan secara grafis atau dalam bentuk laporan yang dapat dicerna dengan lebih mudah oleh pemangku kepentingan.

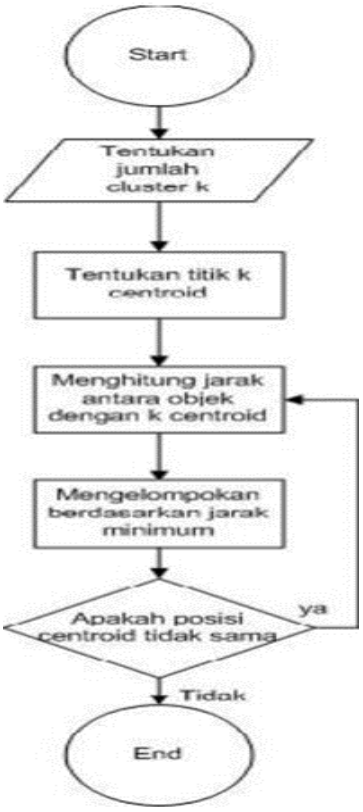
2.2 Metode Clastering

Clastering adalah proses pengelompokan objek serupa ke dalam kelompok yang berbeda atau, lebih tepatnya, partisi data set menjadi subset, sehingga setiap subset memiliki makna yang signifikan. Sebuah claster terdiri dari Kumpulan objek yang mirip satu sama lain dan berbeda dari objek dalam cluster lainnya. Algoritma clastering terdiri dari dua jenis, yaitu hirarkis dan partitional. Algoritma hirarkis menemukan claster secara berurutan dengan claster yang ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada satu waktu. Seperti contoh dalam dunia sosial media, di mana platform mungkin menggunakan klustering untuk mengelompokkan pengguna dengan minat atau perilaku yang serupa. Ini dapat membantu dalam menyajikan konten yang lebih relevan atau rekomendasi teman yang lebih akurat kepada pengguna. (Merliana, Ernawati, & Santoso, 2015)

2.3 Algoritma K-means dan Metode Elbow

Algoritma K-means merupakan algoritma clustering yang kompleks. Algoritma K-means memiliki beberapa parameter yang harus ditentukan , seperti jumlah cluster yang diinginkan dan metode pemilihan titik awal. Parameter tersebut dapat mempengaruhi hasil clustering yang dihasilkan oleh algoritma k-means. K-means digunakan untuk mengelompokan kumpulan data manjadi beberapa kelompok, dengan cara mengelompokkan data memiliki kemiripan berada dalam 1 kelompok dan memisahkan data yang berbeda kedalam kelompok yang berbeda. (firdaus & afrizal, Kmeans Clustering, 2020).

Algoritma K-Means memiliki beberapa kelebihan dan kekurangan dalam implementasinya. Salah satu kelebihannya adalah kemudahan implementasi, serta kemampuannya untuk mengolah data dalam skala besar dengan waktu eksekusi yang relatif cepat. Namun, terdapat kekurangan pada algoritma K-Means, seperti kesulitan dalam menentukan nilai k (jumlah cluster) secara manual, yang dapat mengakibatkan ketidakpastian terkait pilihan nilai k yang optimal. Selain itu, K-Means cenderung kurang efektif ketika dihadapkan dengan data berdimensi tinggi (high-dimensional), dimana performanya dapat menurun. (Trivusi, 2022). Berikut adalah flowchart dari algoritma K-Means, seperti pada Gambar 2.2.



Gambar 2. 2 Flowchart Algoritma K-Means

Langkah-langkah dalam menggunakan algoritma K-Means menurut Merliana, Ernawati, dan Santoso adalah sebagai berikut:

1. Tetapkan nilai *k* sebagai jumlah klaster yang akan di bentuk.
2. Pilih titik pusat klaster awal (centroid) secara acak untuk menentukan nilai k.
3. Hitung jarak antara setiap objek dan centroid di setiap klaster. Untuk mengukur jarak objek ke centroid, gunakan metode Euclidean Distance.

$$Distance(p,q) = \left(\sum_{k}^{n} \mu k \, |PK - qk|r\right) \, 1/r$$

Keterangan :

N = Jumlah data *record* μk = *Bidang* Bobot yang diberikan *pengguna*

K = Urutan *field* data r = 2

4. Alokasikan setiap objek ke centroid yang memiliki jarak terdekat.
5. Lakukan iterasi dan tentukan posisi baru untuk centroid .
6. Ulangi langkah 3 selama posisi centroid baru tidak konsisten.

Metode Elbow merupakan suatu pendekatan yang digunakan untuk mendapatkan informasi dalam menentukan jumlah cluster terbaik. Pendekatan ini melibatkan analisis hasil perbandingan antara jumlah cluster dengan melihat titik di mana penurunan varians menjadi lebih landai, sering disebut sebagai "siku" pada grafik. (Merliana, Ernawati, & Santoso, 2015). Metode Elbow digunakan untuk menentukan jumlah cluster dari dataset sehingga total variasi dalam cluster dapat diminimalkan. Proses penentuan jumlah cluster dimulai dari k=2 dan terus bertambah pada setiap langkah dengan penambahan nilai k sebanyak 1. (Vada & Arief , 2020).

$$SSE = \sum_{K=1}^{K} \sum_{X_{1e}S_K} \|X_i\ C_K\| \frac{2}{2}$$

Dimana

k = jumlah klaster

Xi = nilai atribut data ke-i

Ck = nilai atribut titik pusat

Langkah Langkah metode elbow untuk mennetukan nilai k sebagai berikut:

1. Inisialisasi awal nilai *k*
2. menaikkan nilai *k*
3. menghitung hasil sum of square error dari setiap nilai *k*

4. melihat hasil sum of square error dari nilai k yang turu n secara drastic
5. menetapkan nilai *k* yang berbentuk siku.

2.4 Penerapan K-Means Pada Kasus Perceraian

Penelitian yang dilakukan oleh Uci Dwi Rahayu, Nuke L Chusna, Moch Fachri berjudul "Analisis Kasus Perceraian Pada Pengadilan Negeri Bekasi Dengan Menggunakan Algoritma K-Means Klustering" menghasilkan perhitungan manual jarak Euclidean pada Iterasi Ke tiga. Pada iterasi tersebut, terdapat satu data, yaitu kecamatan Babelan, yang mengalami perpindahan cluster dari data ke-1. Iterasi dilanjutkan dengan melakukan rata-rata setiap cluster untuk menentukan centroid selanjutnya. Hasil perhitungan manual pada Iterasi Ke-3 menunjukkan bahwa hasil cluster tidak berubah, menandakan berakhirnya iterasi karena data tidak berpindah ke cluster lain. Dengan demikian, dapat disimpulkan bahwa terdapat dua kluster dalam kasus perceraian di Pengadilan Negeri Bekasi. Cluster 1 mencakup Jatisampurna, Cikarang Selatan, Bekasi Utara, Pondok Melati, Medan Satria, dan Bekasi Timur. Cluster 2 mencakup Pondokgede, Setu, Bekasi Selatan, Jatiasih, Rawalumbu, Mustika Jaya, Tambun Utara, Tambun Selatan, Serang Baru, Tarumajaya, Cikarang Utara, Babelan, Cibarusah, dan Cibitung. Visualisasi data perceraian dalam bentuk grafik titik menunjukkan dua kluster berdasarkan perbedaan usia pasangan dan lama usia pernikahan. Selanjutnya, karakteristik dua cluster ini dianalisis, dan perbedaan signifikan terlihat pada lama usia pernikahan. Cluster 1 memiliki usia pernikahan yang lebih singkat, sementara Cluster 2 memiliki usia pernikahan yang lebih lama. Dari data ini, dapat

disimpulkan bahwa kasus perceraian di Kota Bekasi, berdasarkan data tahun 2017 hingga 2020, cenderung terkait dengan lama usia pernikahan.

Penelitian yang dilakukan oleh N. Nurhayati, Fahrija Azzahra, Sri Ramadani, Sinta Dwi Hastuti, Eka Irawan berjudul "Analisis Metode Klastering Pada Kasus Penyebab Perceraian Berdasarkan Provinsi Dengan Teknik K-Means" bertujuan untuk menganalisis penyebab perceraian di Indonesia dengan menggunakan metode klastering K-Means. Data dianalisis berdasarkan jumlah penyebab perceraian di setiap provinsi, dan hasilnya di-cluster menjadi dua kelompok menggunakan algoritma K-Means. Cluster tingkat tinggi (C1) terdiri dari Jawa Barat, Jawa Tengah, dan Jawa Timur, sedangkan cluster tingkat rendah (C2) terdiri dari 26 provinsi lainnya. Penelitian ini memberikan informasi penting bagi pemerintah untuk mengembangkan program yang sesuai dengan karakteristik setiap provinsi guna mengurangi tingkat perceraian di Indonesia.

Dengan melakukan studi pendahuluan dan mengamati beberapa penelitian terkait yang relevan mengenai data perceraian serta menerapkan algoritma K-means, penelitian ini memiliki kelebihan dengan menggunakan data yang mencakup atribut jenis perceraian, usia pria (usia p), usia wanita (usia t), faktor penyebab perceraian, dan lokasi tempat KUA kecamatan. Sebagai tambahan, penelitian ini juga menggunakan data yang terkini, yaitu tahun 2022. Selanjutnya, penentuan nilai k dilakukan dengan metode Elbow, yang membantu menentukan jumlah kluster yang optimal. Selain itu, perhitungan kualitas clustering dilakukan dengan menggunakan metode Silhouette Coefficient. Dengan pendekatan ini, penelitian ini diharapkan dapat

memberikan pemahaman yang lebih komprehensif mengenai pola perceraian, serta menawarkan kontribusi dalam penanganan dan analisis data perceraian dengan metode K-means pada tahun 2022.

2.5 Evaluasi Clustering

Dalam mengevaluasi kualitas pengelompokan Dalam proses clustering, kualitas hasil pengelompokan dapat dinilai menggunakan metrik yang disebut Silhouette Coefficient. Untuk menghitung jarak antar data dalam setiap cluster, kita biasanya menggunakan rumus jarak Euclidean. Silhouette Coefficient memberikan penilaian tentang seberapa baik data dikelompokkan dalam sebuah cluster. Metrik ini dihitung untuk setiap cluster maupun untuk keseluruhan cluster dalam hasil algoritma clustering.

Nilai Silhouette untuk keseluruhan data dengan jumlah cluster tertentu, yang kita sebut sebagai sil(k), dihitung berdasarkan rata-rata nilai Silhouette untuk semua cluster. Ini memberikan pandangan umum tentang kualitas clustering dengan jumlah cluster k, membantu kita mengerti seberapa efektif data dikelompokkan oleh algoritma clustering tersebut.**Sumber yang ditentukan tidak valid.**

$$sil(c) = sil(k)\frac{1}{|k|}\sum_{i}^{k} = 1\, sil(Ci)$$

Dimana :

sil (k) :Nilai Silhouette untuk semua cluster.

|k| : banyaknya cluster k

sil (ci) : rata-rata nilai silhouette

BAB III
METODE PENELITIAN

3.1 Analisis Permasalahan

Objek penelitian ini menghadapi permasalahan kompleks terkait identifikasi tingkat perceraian tertinggi di kecamatan pada tahun 2022. Metode manual yang saat ini digunakan untuk sosialisasi dinilai tidak efektif karena mengolah jumlah data yang besar. Dalam mengatasi tantangan ini, clustering diusulkan sebagai solusi untuk mengidentifikasi kecamatan dengan tingkat perceraian paling tinggi.

Penelitian ini diarahkan pada pemahaman bahwa faktor penyebab perceraian dapat bervariasi di setiap individu. Oleh karena itu, penggunaan clustering di harapkan dapat memberikan pemahaman yang lebih mendalam tentang pola perceraian di berbagai kecamatan. Keterbatasan data aktual yang mudah dipahami juga menjadi fokus penelitian, dengan harapan hasil analisis dapat memberikan dasar yang kuat bagi pengadilan untuk mengambil kebijakan yang tegas terkait tingkat perceraian tertinggi.

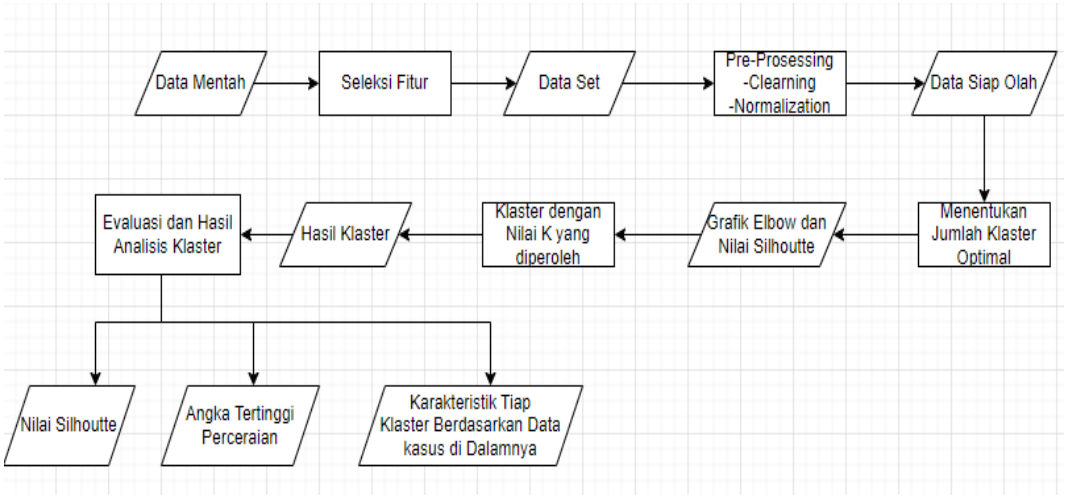
Penelitian sebelumnya belum melibatkan penggunaan clustering untuk menganalisis data perceraian berdasarkan atribut seperti jenis cerai, usia p, usia t, faktor penyebab, dan tempat pernikahan (Kantor Urusan Agama/KUA). Selain itu, untuk mengidentifikasi kecamatan dengan tingkat perceraian tertinggi berdasarkan karakteristiknya. Akibatnya, Pengadilan menghadapi kesulitan dalam mengambil kebijakan yang berdasarkan data perceraian di setiap kecamatan.

Dalam konteks ini, penting untuk mengembangkan pendekatan analisis yang lebih holistik dan dinamis untuk memberikan pemahaman yang lebih baik tentang pola perceraian, sehingga kebijakan yang diambil dapat lebih tepat sasaran dan responsif terhadap kebutuhan masyarakat di berbagai kecamatan.

3.2 Solusi yang Diusulkan

Solusi yang diajukan untuk mengatasi permasalahan tingkat perceraian di wilayah Kabupaten Lamongan adalah dengan melakukan klasterisasi data perceraian menggunakan metode K-means. Tujuan dari klasterisasi data perceraian ini adalah untuk mempermudah analisis terhadap data yang terkait dengan tingkat perceraian tertinggi di Kabupaten Lamongan berdasarkan karakteristik, sehingga dapat diambil kebijakan yang tepat untuk menanggulangi permasalahan tersebut. Langkah awal dalam implementasi metode K-means adalah menentukan jumlah cluster terbaik. Penentuan ini dilakukan dengan menggunakan metode Elbow, di mana kita melihat siku yang terbentuk pada grafik, memberikan indikasi jumlah klaster yang optimal. Selanjutnya, kualitas klaster yang terbentuk dievaluasi menggunakan metode Silhouette Coefficient. Metode ini digunakan untuk mengukur sejauh mana objek-objek dalam satu klaster seragam dan seberapa baik klaster tersebut terpisah satu sama lain. Dengan mengadopsi metode K-means dan menerapkan langkah-langkah evaluasi tersebut, diharapkan Pengadilan Kabupaten Lamongan dapat mendapatkan pemahaman yang lebih baik tentang pola perceraian di berbagai wilayah. Informasi ini di harapkan dapat menjadi dasar yang kokoh untuk merumuskan kebijakan yang efektif

dalam menangani permasalahan tingkat perceraian yang dihadapi. Dalam penelitian ini terdapat beberapa tahapan penelitian seperti pada Gambar 3.1.



Gambar 3. 1 Flowchart Penelitian

3.2.1 **Seleksi Fitur**

Dalam penelitian ini, seleksi fitur merujuk pada pemilihan atribut yang digunakan sebagai dasar untuk melakukan klasterisasi. Studi ini memilih 5 atribut sebagai fokus penelitian, yaitu Jenis Cerai, Umur p, Umur t, Faktor Penyebab, dan Tempat Pernikahan (KUA). Perceraian adalah kondisi yang menandakan berakhirnya hukum perkawinan, sehingga kedua belah pihak tidak lagi dianggap sebagai suami istri dan tidak menjalani kehidupan bersama dalam suatu rumah tangga.

Rata-rata perbedaan antar jenis cerai, seperti cerai talak dan cerai gugat, menjadi sorotan dalam penelitian ini. Umur p dianggap sebagai faktor penting yang dapat mengajukan gugatan dengan kepentingan yang memadai, sementara umur t mengacu pada orang yang digugat oleh Penggugat. Faktor penyebab perceraian beragam antar individu, begitu juga

dengan tempat pernikahan atau KUA, yang berbeda di setiap dari 29 kecamatan di Kabupaten Lamongan. Atribut-atribut ini juga dijadikan standar untuk mengukur tingkat perceraian tertinggi di wilayah Kabupaten Lamongan oleh pengadilan.

3.2.2 **Preprocessing**

Segala bentuk pemrosesan data mentah diperlukan agar data tersebut siap untuk diolah pada tahapan analisis data berikutnya. Proses data preprocessing bertujuan mengubah format data sehingga menjadi lebih mudah dan efektif dalam proses data mining (Vikra , 2023).

1. Cleansing adalah suatu proses yang melibatkan perbaikan atau penghapusan data yang mungkin salah, rusak, memiliki format yang tidak sesuai, duplikat, atau tidak lengkap dalam suatu kumpulan data. Ketika menggabungkan beberapa sumber data, terdapat berbagai peluang bagi duplikasi data atau kesalahan pelabelan. Jika data tidak benar, maka hasil dan algoritme yang digunakan tidak dapat diandalkan, meskipun mungkin terlihat benar. Tidak ada satu cara mutlak untuk menentukan langkah-langkah pasti dalam proses pembersihan data karena proses ini bervariasi antara satu kumpulan data dan kumpulan data lainnya. Pada tahap ini, tidak dilakukan penghapusan data karena tidak ada data yang kosong dalam dataset yang diperoleh.

2. Normalisasi merupakan proses penyesuaian nilai-nilai atribut agar memiliki skala yang seragam, sehingga tidak ada atribut yang mendominasi yang lainnya. Dengan Atribut numerik dapat diskalakan ke dalam rentang

yang lebih kecil, misalnya 1,10. Teknik normalisasi seperti metode Min-max Normalization, Min-Max scaling dan Z-score normalization dapat diterapkan untuk mencapai tujuan ini pada tingkat klaster atau kelompok data tertentu.

a. Min-max Normalization

Memetakan nilai dari sebuah atribut ke dalam suatu rentang dilakukan dengan menggunakan rumus perhitungan berikut:

Min : =MIN(Data Awal – Data Akhir)

Max : =MAX(Data Awal – Data Akhir)

b. Min-Max scaling

Min-Max Scaling beroperasi dengan cara menyesuaikan data ke dalam rentang tertentu, yang umumnya adalah dari nilai minimum hingga nilai maksimum. Rentang yang biasa digunakan adalah 0 hingga 1. Dengan menggunakan rumus:

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

c. Z-score normalization

Mengacu pada proses normalisasi di mana setiap nilai dalam kumpulan data disesuaikan sehingga rata-rata dari semua nilai menjadi 0, dan deviasi standarnya menjadi 1. Rumus yang digunakan:

(Data ke-1 – Min.All) / (Max.All – Min.All).

Dalam penelitian ini, dilakukan normalisasi data menggunakan metode min-max normalization dalam bahasa pemrograman Python. Normalisasi bermanfaat untuk mengubah data yang memiliki rentang nilai yang berbeda

menjadi data dengan rentang nilai yang seragam, sehingga mempermudah perbandingan dan pengolahan data.

3.2.3 **Menentukan Jumlah Cluster Optimal**

Clustering menggunakan algoritma K-means melibatkan penentuan nilai optimal untuk k, yang dapat diperoleh melalui metode Elbow. Yang di hasilkan dari perhitungan nilai silhouette coefficient. Langkah berikutnya adalah membuat objek n_clusters yang berisi pengaturan KMeans dengan jangkauan dari 2 hingga 11. Kemudian, siapkan sebuah list kosong bernama silhoutte_scores untuk menyimpan skor perhitungan silhouette coefficient untuk setiap nilai k. Untuk melakukan ini, kita dapat menggunakan teknik K-means++ yang memilih titik centroid awal dengan cara yang lebih efisien, yaitu dengan pengambilan sampel secara acak. Setelah inisialisasi centroid dilakukan, perhitungan jarak setiap data ke centroidnya dapat dihitung. Data yang di-cluster harus dianalisis untuk menentukan seberapa besar pengaruh setiap atribut terhadap atribut lainnya. Pemberian label pada setiap centroid penting untuk memahami perbedaan antar clustering yang dihasilkan. Ini membantu dalam interpretasi hasil dan memberikan pemahaman tentang karakteristik setiap cluster. langkah terakhir adalah menampilkan hasil clustering yang telah dilakukan. Dengan visualisasi yang baik, kita dapat dengan jelas melihat pembagian data ke dalam cluster yang sesuai, membantu interpretasi dan pengambilan keputusan lebih lanjut.

3.2.4 Evaluasi dan Hasil Analisis Cluster

Evaluasi cluster menggunakan metode Silhoutte Coefficient bertujuan untuk melihat kualitas dan kekuatan cluster serta mengukur seberapa baik suatu objek ditempatkan dalam suatu cluster. Tahapan yang dilakukan dalam perhitungan nilai Silhoutte Coefficient yaitu dengan menghitung rata-rata jarak dari suatu data dengan data lain yang berada dalam satu cluster. Selanjutnya menghitung rata-rata jarak data tersebut dengan semua data di cluster lain dan diambil nilai terkecilnya. Untuk menghitung nilai Silouhette Coefficient diharuskan mengimport silhouette_score dari library sklearn. Berdasarkan karakteristik klaster yang diperoleh, dapat lebih mudah merumuskan kebijakan dalam penanggulangan perceraian di tiap kecamatan.

3.2 Rancangan Eksperimen

3.3.1 Peralatan

Peralatan yang digunakan untuk mengimplementasikan penelitian ini terdapat bebrapa kompoenen sebagai berikut:

1. **Perangkat Keras (Hardware)**

   a. Laptop dengan processor Intel(R) Core(TM) i5-6200U.

   b. RAM 8,00 GB.

   c. 64-bit operating system.

2. **Perangkat Lunak (Software)**

   a. Python version 3.12.0 bahasa pemrograman yang digunakan untuk mengimplementasikan program.

b. Library Pandas, Numpy, Matplotlib, Seaborn dan Scikit-learn, digunakan untuk mengimplementasikan dari algoritma K-means clustering.

c. Collab Notebook, digunakan untuk mengeksekusi program berbahasa python.

d. Sistem Operasi Windows 10

e. Microsoft Excel, digunakan untuk melakukan pengolahan data sekunder yang diperoleh dari Pengadilan.

3.3.2 Kebutuhan Data

Data yang digunakan untuk pengujian berasal dari hasil wawancara dan observasi di Pengadilan Agama Lamongan berdasrkan rekomendasi dari hakim kemudian mendapatkan data dari Panitera, yang mencakup informasi mengenai kasus perceraian yang terjadi pada tahun 2020 - 2022. Dataset ini terdiri dari lima atribut, yaitu jenis perceraian, umur p (penggugat/pemohon) merujuk kepada pihak yang mengajukan permohonan atau gugatan, umur t (tergugat/termohon) merujuk pada pihak yang dihadapkan dengan gugatan atau permohonan, faktor penyebab perceraian, dan KUA tempat nikah (kecamatan). Jumlah total rekord dalam dataset pada tahun 2020 (2486), 2021 (2689), 2022 (2608) sesuai dengan jumlah kasus perceraian. Data tersebut disimpan dalam satu file XLSX, dan proses pengolahan data mentah dilakukan menggunakan Microsoft Excel.

BAB IV
PENGUJIAN DAN HASIL

4.1 Penerapan Metode

4.2.1 Seleksi Fitur

Tahap seleksi fitur dilakukan dengan menggabungkan atribut yang terdapat dalam setiap file Excel yang diperoleh dari observasi dan wawancara di Pengadilan Agama Lamongan. Penelitian ini memanfaatkan Lima atribut dengan mengambil data dari tahun 2022, yang menghasilkan satu file Excel dengan data yang siap untuk diolah. Totalnya, ada satu file Excel. Berikut merupakan sample data berdasarkan atribut, seperti pada Tabel 4.1.

Tabel 4. 1 Sample Data Mentah Penceraian 2020

REKAPITULASI TERJADINYA PERCERAIAN					
PENGADILAN AGAMA LAMONGAN					
TAHUN 2020					
no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
1	Cerai Gugat	30	31	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Sambeng, Kabupaten Lamongan
2	Cerai Gugat	27	33	Ekonomi	Kantor Urusan Agama Kecamatan Bluluk, Kabupaten Lamongan
3	Cerai Gugat	38	51	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Paciran, Kabupaten Lamongan
4	Cerai Gugat	26	27	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama (KUA) Kecamatan Solokuro Kabupaten Lamongan
5	Cerai Gugat	26	36	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Brondong, Kabupaten Lamongan
6	Cerai Gugat	43	52	Ekonomi	Kantor Urusan Agama Kecamatan Kedungpring, Kabupaten Lamongan
7	Cerai Gugat	21	23	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Kembangbahu, Kabupaten Lamongan
8	Cerai Gugat	64	63	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Karangbinangun, Kabupaten Lamonga
9	Cerai Gugat	35	43	Ekonomi	Kantor Urusan Agama Kecamatan Lamongan, Kabupaten Lamongan
10	Cerai Gugat	38	40	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Babat, Kabupaten Lamongan
11	Cerai Gugat	24	32	Ekonomi	Kantor Urusan Agama Kecamatan Maduran, Kabupaten Lamongan
12	Cerai Gugat	24	29	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Sambeng, Kabupaten Lamongan
13	Cerai Gugat	52	60	Ekonomi	Kantor Urusan Agama Kecamatan Bluluk, Kabupaten Lamongan
14	Cerai Talak	30	30	Ekonomi	Kantor Urusan Agama Kecamatan Deket, Kabupaten Lamongan
15	Cerai Talak	36	29	Meninggalkan Salah Satu Pihak	Kantor Urusan Agama Kecamatan Laren, Kabupaten Lamongan
16	Cerai Talak	37	37	Meninggalkan Salah Satu Pihak	Kantor Urusan Agama Kecamatan Solokuro, Kabupaten Lamongan
17	Cerai Talak	39	40	Ekonomi	Kantor Urusan Agama Kec. Babat, Kab. Lamongan

Tabel 4. 2 Sample Data Mentah Penceraian 2021

REKAPITULASI TERJADINYA PERCERAIAN					
PENGADILAN AGAMA LAMONGAN					
TAHUN 2021					
no	jenis_cera	umur_p	umur_	faktor_perceraian	kua_tempat_nikah
1	Cerai Talak	26	25	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Kembang bahu Kabupaten Lamongan
2	Cerai Talak	48	40	Ekonomi	KUA Kecamatan Nanga Pinoh, Kabupaten Sintang
3	Cerai Talak	50	41	Ekonomi	KUA Kecamatan Modo, Kabupaten Lamongan
4	Cerai Talak	42	38	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Duduksampeyan, Kabupaten Gresik
5	Cerai Talak	27	23	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Kedungpring, Kabupaten Lamongan
6	Cerai Gugat	19	22	Ekonomi	KUA Kecamatan Kedungpring, Kabupaten Lamongan
7	Cerai Gugat	51	53	Judi	KUA Kecamatan Babat, Kabupaten Lamongan
8	Cerai Gugat	27	29	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Brondong Kabupaten Lamongan
9	Cerai Gugat	41	50	Ekonomi	KUA Kecamatan Sukodadi, Kabupaten Lamongan
10	Cerai Gugat	35	2020	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Paciran Kabupaten Lamongan
11	Cerai Gugat	29	26	Ekonomi	Kantor Urusan Agama Kecamatan Babat, Kabupaten Lamongan
12	Cerai Gugat	27	31	Ekonomi	Kantor Urusan Agama (KUA) Kecamatan Mantup, Kabupaten Lamongan
13	Cerai Gugat	38	39	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Ngimbang, Kabupaten Lamongan
14	Cerai Gugat	2020	38	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Tikung, Kabupaten Lamongan
15	Cerai Gugat	38	32	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Kembangbahu, Kabupaten Lamongan
16	Cerai Gugat	23	31	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sugio, Kabupaten Lamongan
17	Cerai Gugat	25	35	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Babat, Kabupaten Lamongan

Tabel 4. 3 Sample Data Mentah Penceraian 2022

REKAPITULASI TERJADINYA PERCERAIAN
PENGADILAN AGAMA LAMONGAN
TAHUN 2022

no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
1	Cerai Gugat	24	27	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Turi Kabupaten Lamongan
2	Cerai Gugat	32	36	Ekonomi	KUA Kecamatan Ngimbang, Kabupaten Lamongan
3	Cerai Gugat	31	41	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Babat, Kabupaten Lamongan
4	Cerai Gugat	27	24	Meninggalkan Salah Satu Pihak	KUA Kecamatan Kalitengah, Kabupaten Lamongan
5	Cerai Gugat	38	36	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Babat, Kabupaten Lamongan
6	Cerai Gugat	22	24	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sekaran Kabupaten Lamongan
7	Cerai Gugat	33	43	Ekonomi	KUA Kecamatan Babat, Kabupaten Lamongan
8	Cerai Gugat	45	49	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Karangbinangun, Kabupaten Lamongan
9	Cerai Gugat	54	59	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Mantup, Kabupaten Lamongan
10	Cerai Gugat	36	38	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sekaran, Kabupaten Lamongan
11	Cerai Gugat	24	24	Ekonomi	KUA Kecamatan Sarirejo, Kabupaten Lamongan
12	Cerai Gugat	41	47	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Karanggeneng, Kabupaten Lamongan
13	Cerai Gugat	30	37	Ekonomi	KUA Kecamatan Sambeng, Kabupaten Lamongan
14	Cerai Talak	30	24	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Karanggeneng, Kabupaten Lamongan
15	Cerai Gugat	39	39	Meninggalkan Salah Satu Pihak	KUA Kecamatan Kembangbahu, Kabupaten Lamongan
16	Cerai Gugat	48	60	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sugio, Kabupaten Lamongan
17	Cerai Gugat	28	28	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Lamongan, Kabupaten Lamongan
18	Cerai Gugat	39	42	Ekonomi	KUA Kecamatan Lamongan, Kabupaten Lamongan
19	Cerai Gugat	30	31	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Modo, Kabupaten Lamongan

4.2.2 **Pre-Processing**

Dalam tahap pre-processing, data mentah diproses untuk persiapan analisis lanjutan. Tahapan ini melibatkan beberapa langkah penting, seperti Cleansing, transformasi dan normalisasi data. Penjelasan mendetail mengenai proses ini akan disajikan pada bagian berikutnya dari penelitian ini.

4.2.3 **Cleaning Data**

Tahap cleaning data bertujuan untuk mengecek data bagian baris ataupun kolom yang kosong. Pada penelitian ini tidak ada data yang kosong.

```
missing_rows = data.index[data.isnull().any(axis=1)].tolist()
print("\nData dengan missing value berada pada baris ke:", missing_rows)
```

Gambar 4. 1 Source Code Cleansing

Berikut merupakan hasil dari tahapan cleaning data 2020 - 2022

```
Data dengan missing value berada pada baris ke: []
```

Gambar 4. 2 Hasil Cleansing

4.2.4 Transformasi Data

Pada tahapan transformasi data bertujuan mengubah variabel kategori menjadi bentuk numerik untuk digunakan dalam analisis serta proses lanjutan. Transformasi data dilakukan dengan menambahkan kolom baru dengan nama angka_kua_tempat_nikah pada file data mentah dalam format Excel, yang kemudian diisi dengan angka yang sesuai dengan kondisi kecamatan tempat pelaksanaan pernikahan. Untuk jumlah angka mencapai 29 sesuai dengan jumlah Kecamatan di Kabupaten Lamongan. Berikut merupakan nilai angka sesuai dengan kua tempat nikah.

1. Import Data

   Import data dilakukan dengan mengambil data dalam format Excel yang telah diunggah ke Google Drive.

```
import pandas as pd  # Mengimpor pustaka pandas dengan alias 'pd'
data = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/data malinda skripsi 2020.xlsx')
data
```

Gambar 4. 3 Source Code Import Data 2020

```
import pandas as pd  # Mengimpor pustaka pandas dengan alias 'pd'
data = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/data malinda skripsi 2021.xlsx')
data
```

Gambar 4. 4 Souce Code Import Data 2021

```
import pandas as pd  # Mengimpor pustaka pandas dengan alias 'pd'
data = pd.read_excel('/content/drive/MyDrive/Colab Notebooks/data malinda skripsi 2022.xlsx')
data
```

Gambar 4. 5 Source Code Import Data 2022

Berikut ini adalah tampilan data yang berhasil diimport tahun 2020 - 2022, dapat dilihat pada gambar berikut :

	no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	1	Cerai Gugat	30	31	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Sambeng, Kabupat...
1	2	Cerai Gugat	27	33	Ekonomi	Kantor Urusan Agama Kecamatan Bluluk, Kabupate...
2	3	Cerai Gugat	38	51	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Paciran, Kabupat...
3	4	Cerai Gugat	26	27	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama (KUA) Kecamatan Solokuro K...
4	5	Cerai Gugat	26	36	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Brondong, Kabupa...
...	...	...	...	...	...	...
2480	2481	Cerai Gugat	63	50	Ekonomi	KUA KecamatanSekaran, Kabupaten Lamongan
2481	2482	Cerai Gugat	52	53	Zina	KUA Kecamatan Modo, Kabupaten Lamongan
2482	2483	Cerai Gugat	27	44	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Modo, Kabupaten Lamongan
2483	2484	Cerai Gugat	41	38	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Mantup, Kabupaten Lamongan
2484	2485	Cerai Gugat	39	33	Ekonomi	KUA Kecamatan Ngimbang, Kabupaten Lamongan

2485 rows × 6 columns

Gambar 4. 6 Hasil Import Data Set 2020

	no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	1	Cerai Talak	26	25	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Kembang bahu Kab...
1	2	Cerai Talak	48	40	Ekonomi	KUA Kecamatan Nanga Pinoh, Kabupaten Sintang
2	3	Cerai Talak	50	41	Ekonomi	KUA Kecamatan Modo, Kabupaten Lamongan
3	4	Cerai Talak	42	38	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Duduksampeyan, K...
4	5	Cerai Talak	27	23	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Kedungpring, Kabupaten Lamongan
...	...	...	...	...	...	...
2678	2679	Cerai Gugat	35	40	Meninggalkan Salah Satu Pihak	KUA Kecamatan Karangbinangunj, Kabupaten Lamongan
2679	2680	Cerai Gugat	35	47	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sukorame, Kabupaten Lamongan
2680	2681	Cerai Gugat	26	29	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Bluluk, Kabupaten Lamongan
2681	2682	Cerai Gugat	31	32	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Mantup, Kabupaten Lamongan
2682	2683	Cerai Gugat	45	42	Ekonomi	KUA Kecamatan Modo, Kabupaten Lamongan

2683 rows × 6 columns

Gambar 4. 7 Hasil Import Data Set 2021

	no	jenis_cerai	angka jenis	umur_p	umur_t	faktor_perceraian	angka_faktor	kua_tempat_nikah	angka_kua_tempat_nikah
0	1	Cerai Gugat	1	24	27	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Turi Kabupaten Lamongan	27
1	2	Cerai Gugat	1	32	36	Ekonomi	2	KUA Kecamatan Ngimbang, Kabupaten Lamongan	16
2	3	Cerai Gugat	1	31	41	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Babat, Kabupaten Lamongan	1
3	4	Cerai Gugat	1	27	24	Meninggalkan Salah Satu Pihak	3	KUA Kecamatan Kalitengah, Kabupaten Lamongan	6
4	5	Cerai Gugat	1	38	36	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Babat, Kabupaten Lamongan	1
...	...	...	...	...	...	...	...	...	...
2602	2603	Cerai Gugat	1	26	36	Ekonomi	2	KUA Kecamatan Sugio, Kabupaten Lamongan	23
2603	2604	Cerai Gugat	1	42	49	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Kedungpring, Kabupaten Lamongan	9

Gambar 4. 8 Hasil Import Data Set 2022

2. Mengambil value untuk tranformasi data

Mengonversi data menjadi format numerik pada kolom 'jenis_cerai', 'umur_penggugat', 'umur_tergugat', 'faktor_perceraian', 'kua_tempat_nikah' dilakukan dengan menggunakan LabelEncoder. Label encoder merupakan metode yang digunakan untuk mengonversi nilai kategorikal menjadi nilai numerik secara otomatis.

```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
cols_to_encode = ['jenis_cerai', 'faktor_perceraian', 'kua_tempat_nikah']
data_selec[cols_to_encode] = data_selec[cols_to_encode].apply(lambda col: label_encoder.fit_transform(col))
data_selec.to_excel('/content/drive/MyDrive/Colab Notebooks/data_sebelum di normalisasi.xlsx')
print(data_selec)
```

Gambar 4. 9 Source Code Transformasi Data

Berikut merupakan hasil dari transformasi data menggunakan LabelEncoder tahun 2020 – 2022.

Berikut merupakan hasil dari tranfomasi data tahun 2020 seperti pada gambar berikut:

	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	0	30	31	10	363
1	0	27	33	2	177
2	0	38	51	10	331
3	0	26	27	10	113
4	0	26	36	10	183
...	...	...	...	...	...
2480	0	63	50	2	65
2481	0	52	53	12	39
2482	0	27	44	10	39
2483	0	41	38	10	36
2484	0	39	33	2	40

[2485 rows x 5 columns]

Gambar 4. 10 Hasil Transformasi Data 2020

Berikut merupakan hasil dari tranfomasi data tahun 2021 seperti pada gambar berikut:

	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	1	26	25	10	458
1	1	48	40	2	209
2	1	50	41	2	207
3	1	42	38	10	440
4	1	27	23	10	160
...	...	...	...	...	...
2678	0	35	40	8	145
2679	0	35	47	10	295
2680	0	26	29	10	74
2681	0	31	32	10	195
2682	0	45	42	2	207

[2683 rows x 5 columns]

Gambar 4. 11 Hasil Transformasi Data 2021

Berikut merupakan hasil dari tranfomasi data tahun 2022 seperti pada gambar berikut:

	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	0	24	27	10	279
1	0	32	36	2	192
2	0	31	41	10	31
3	0	27	24	8	102
4	0	38	36	10	31
...	...	...	...	...	...
2602	0	26	36	2	247
2603	0	42	49	10	121
2604	0	19	25	2	125
2605	0	22	36	10	194
2606	0	27	29	2	196

[2607 rows x 5 columns]

Gambar 4. 12 Hasil Transformasi Data 2022

4.2.5 **Normalisasi Data**

1. Perhitungan normalisasi exsel

Langkah awal untuk menemukan nilai minimum dan maksimum dari setiap kolom seperti berikut mencari nilai terkecil dalam rentang sel dari B2 hingga B2486.

=MIN(B2:B2486)

Min merupakan hasil minimum b2 merupakan baris kolom di bagi baris kolom dengan jumlah data.

=MAX(B2:B2684)

Max merupakan hasil minimum b2 merupakan kolom baris di bagi bari kolom dengan jumlah data, mencari nilai terbesar dalam rentang sel dari B2 hingga B2486

Kemudian hitung range (rentang) untuk setiap kolom seperti berikut:

=MAX(B2:B2684) - MIN(B2:B2684)

Maxsimum b2 merupakan kolom di bagi b2 dengan jumlah data di kurangi dengan minimun kolom dibagi dengan kolom dengan jumlah data.

Implementasi di Spreadsheet sebagai berikut

=(B2 - MIN(B2:B2486)) / 1

(B2) Fungsi MIN digunakan untuk mencari nilai minimum dari rentang sel B2 sampai B2486. Nilai minimum ini berfungsi sebagai titik acuan untuk normalisasi.

B2 - MIN(B2) Ini menghitung selisih antara nilai pada sel B2 dengan nilai minimum dari seluruh rentang. Selisih ini menunjukkan seberapa jauh nilai B2 dari nilai minimum. / 1 Ini adalah pembagian hasil selisih dengan 1. Dalam konteks

normalisasi, pembagian dengan 1 tidak mengubah nilai hasil perhitungan selisih di atas.

=(C2 - MIN(C2:C2486)) /83

C2: Ini adalah nilai data dalam kolom C yang ingin dinormalisasi.MIN(C2) Fungsi MIN digunakan untuk mencari nilai minimum dari rentang sel C2 sampai C2486. Nilai minimum ini berfungsi sebagai titik acuan untuk normalisasi.

C2 - MIN(C2) Ini menghitung selisih antara nilai pada sel C2 dengan nilai minimum dari seluruh rentang. Selisih ini menunjukkan seberapa jauh nilai C2 dari nilai minimum. / 83 Hasil selisih tadi kemudian dibagi dengan 83.

=(D2 - MIN(D2:D2486)) / 2020

D2: Ini adalah nilai yang ingin dinormalisasi. Misalkan ini adalah nilai di sel D2. MIN(D2:D2486): Fungsi MIN digunakan untuk mencari nilai minimum dari rentang data yang mencakup dari sel D2 hingga D2486. Ini berarti mencari nilai terkecil dalam keseluruhan rentang data tersebut. D2 - MIN(D2:D2486): Setelah menemukan nilai minimum dalam rentang data tersebut, nilai ini kemudian dikurangi dari nilai di sel D2. Ini adalah langkah untuk menggeser nilai di D2 sehingga menjadi relatif terhadap nilai minimum dalam rentang data. (D2 - MIN(D2:D2486)) / 2020: Hasil pengurangan tersebut kemudian dibagi dengan 2020. Pembagian ini bertujuan untuk mengubah skala nilai yang sudah di-offset tersebut. Biasanya, angka 2020 dalam konteks ini mungkin digunakan untuk mengatur rentang data agar lebih sesuai atau normalisasi ke dalam rentang tertentu, tergantung pada konteks data.

=(E2 - MIN(E2:$E$2486)) / 12

untuk menggunakan tanda dolar ($) untuk membuat referensi sel absolut untuk nilai minimum dan maksimum, sehingga mereka tidak berubah saat rumus disalin ke sel lain.

=(F2 - MIN(F2:$F$2486)) /463

F2: Ini adalah nilai data dalam kolom F yang ingin dinormalisasi.

MIN(F2:$F$2486): Fungsi MIN digunakan untuk mencari nilai minimum dari rentang sel F2 sampai $F$2486. Penulisan $F$2486 dengan tanda dolar ($) menunjukkan bahwa referensi sel ini bersifat absolut, artinya tidak akan berubah saat rumus disalin ke sel lain.

F2 - MIN(F2:$F$2486): Ini menghitung selisih antara nilai pada sel F2 dengan nilai minimum dari rentang yang disebutkan. Selisih ini menunjukkan seberapa jauh nilai F2 dari nilai minimum.

Hasil perhitungan normaliasi data tahun 2020 – 2022 sebagai berikut:

jenis cerai	umur p	umur t	faktor penyebab	kua tempat nikah
0	0.361446	0.015347	0.833333333	0.784017279
0	0.325301	0.016337	0.166666667	0.382289417
0	0.457831	0.025248	0.833333333	0.714902808
0	0.313253	0.013366	0.833333333	0.244060475
0	0.313253	0.017822	0.833333333	0.39524838
0	0.518072	0.025743	0.166666667	0.537796976
0	0.253012	0.011386	0.833333333	0.55075594
0	0.771084	0.031188	0.833333333	0.505399568

Gambar 4. 13 Hasil Normalisasi Data Tahun 2020

jenis_cera	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
1	0.012871	0.009921	0.833333333	0.860902256
1	0.023762	0.017361	0.166666667	0.392857143
1	0.024752	0.017857	0.166666667	0.389097744
1	0.020792	0.016369	0.833333333	0.827067669
1	0.013366	0.008929	0.833333333	0.30075188
0	0.009406	0.008433	0.166666667	0.30075188
0	0.025248	0.02381	0.25	0.103383459
0	0.013366	0.011905	0.833333333	0.806390977

Gambar 4. 14 Hasil Normalisasi Data Tahun 2021

jenis cera	umur p	umur t	faktor perceraian	kua tempat nikah
0	0.114754	0.005484	0.833333333	0.871875
0	0.245902	0.00997	0.166666667	0.6
0	0.229508	0.012463	0.833333333	0.096875
0	0.163934	0.003988	0.666666667	0.31875
0	0.344262	0.00997	0.833333333	0.096875
0	0.081967	0.003988	0.833333333	0.715625
0	0.262295	0.01346	0.166666667	0.096875
0	0.459016	0.016451	0.833333333	0.346875

Gambar 4. 15 Hasil Normalisasi Data Tahun 2022

2. Perhitungan normalisasi data dari collab

Normalisasi data adalah langkah penting dalam mengubah data ke dalam format numerik dengan skala yang lebih terbatas, umumnya dari 0,0 hingga 1,0. Dalam konteks penelitian ini, normalisasi data dilakukan melalui penerapan metode min-max normalization yang diimplementasikan dalam platform Collab.

Pada baris pertama pada gambar 4.7 menjelaskan tentang membuat objek dari kelas Standard MinMax Scaler untuk melakukan pensekalaan fitur. Pada baris selanjutnya menggunakan metode fit_transform()dari objek Standard MinMax Scaler untuk melakukan penskalaan fitur pada data yang telah di pilih

sebelumnya (‘data_selec’). Proses fitting dan tranformasi dilakukan secara bersamaan.

```
from sklearn.preprocessing import MinMaxScaler
scaler =MinMaxScaler()
df_normalized = scaler.fit_transform(data_selec)
df_normalized = pd.DataFrame(df_normalized, columns=data_selec.columns)
df_normalized.to_excel('/content/drive/MyDrive/Colab Notebooks/data_setelah di normalisasi.xlsx')
print(df_normalized)
```

Gambar 4. 16 Source Code Normalisasi Data

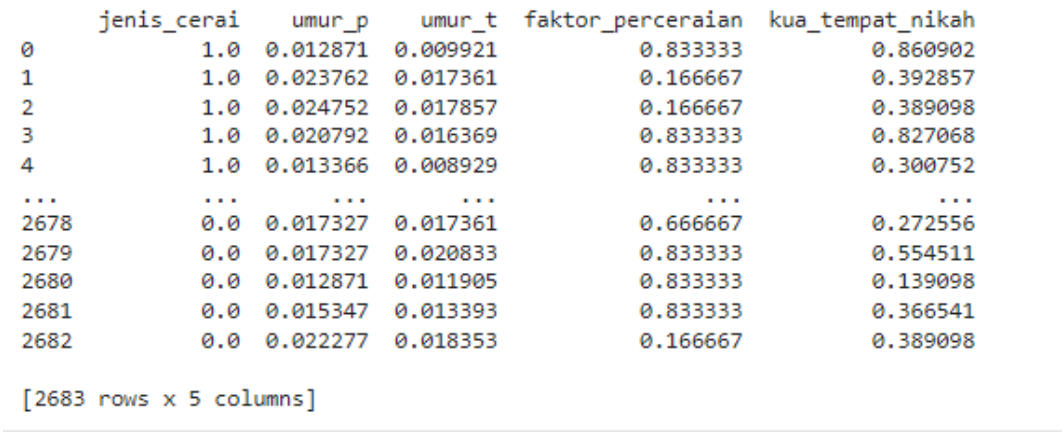
Berikut merupakan hasil normalisasi data tahun 2020 menggunakan min-max normalization seperti pada gambar

	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	0.0	0.361446	0.015347	0.833333	0.784017
1	0.0	0.325301	0.016337	0.166667	0.382289
2	0.0	0.457831	0.025248	0.833333	0.714903
3	0.0	0.313253	0.013366	0.833333	0.244060
4	0.0	0.313253	0.017822	0.833333	0.395248
...	...	...	...	...	...
2480	0.0	0.759036	0.024752	0.166667	0.140389
2481	0.0	0.626506	0.026238	1.000000	0.084233
2482	0.0	0.325301	0.021782	0.833333	0.084233
2483	0.0	0.493976	0.018812	0.833333	0.077754
2484	0.0	0.469880	0.016337	0.166667	0.086393

[2485 rows x 5 columns]

Gambar 4. 17 Hasil Normalisasi Data 2020

Berikut merupakan hasil normalisasi data tahun 2021 menggunakan min-max normalization seperti pada gambar

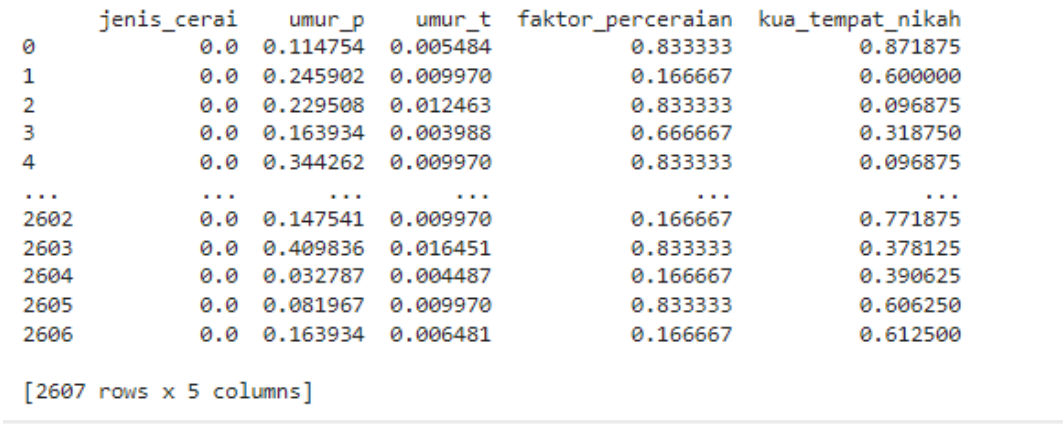


	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	1.0	0.012871	0.009921	0.833333	0.860902
1	1.0	0.023762	0.017361	0.166667	0.392857
2	1.0	0.024752	0.017857	0.166667	0.389098
3	1.0	0.020792	0.016369	0.833333	0.827068
4	1.0	0.013366	0.008929	0.833333	0.300752
...	...	...	...	...	...
2678	0.0	0.017327	0.017361	0.666667	0.272556
2679	0.0	0.017327	0.020833	0.833333	0.554511
2680	0.0	0.012871	0.011905	0.833333	0.139098
2681	0.0	0.015347	0.013393	0.833333	0.366541
2682	0.0	0.022277	0.018353	0.166667	0.389098

[2683 rows x 5 columns]

Gambar 4. 18 Hasil Normalisasi Data 2021

Berikut merupakan hasil normalisasi data tahun 2022 menggunakan min-max normalization seperti pada gambar



	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah
0	0.0	0.114754	0.005484	0.833333	0.871875
1	0.0	0.245902	0.009970	0.166667	0.600000
2	0.0	0.229508	0.012463	0.833333	0.096875
3	0.0	0.163934	0.003988	0.666667	0.318750
4	0.0	0.344262	0.009970	0.833333	0.096875
...	...	...	...	...	...
2602	0.0	0.147541	0.009970	0.166667	0.771875
2603	0.0	0.409836	0.016451	0.833333	0.378125
2604	0.0	0.032787	0.004487	0.166667	0.390625
2605	0.0	0.081967	0.009970	0.833333	0.606250
2606	0.0	0.163934	0.006481	0.166667	0.612500

[2607 rows x 5 columns]

Gambar 4. 19 Hasil Normalisasi Data 2022

4.2.6 **Menentukan Jumlah Cluster Optimal**

Penentuan jumlah cluster optimal menggunakan metode elbow melibatkan perhitungan nilai Silhouette Coefficient. Kombinasi kedua metode tersebut bertujuan untuk mendapatkan cluster yang optimal dan berkualitas. Berikut adalah

langkah-langkah yang diambil untuk menentukan nilai k yang efektif dengan silhoutte coeficient dan metode elbow. Berikut merupakan hasil kodingan

```
inertias = []
silhouette_avg
# Menghitung inertia dan Silhouette score untuk setiap jumlah cluster
for k in range(2, 11)
kmeans = KMeans(n_clusters= k, random_state=0,  init = 'k-means++',max_iter=100)
kmeans.fit(df_normalized)
inertias.append(kmeans.inertia_)
silhouette_avg.append(silhouette_score(df_normalized, kmeans.labels_))
# Plotting kedua metrik dalam satu plot
# Plotting inertia
# Plotting inertia
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), inertias, marker='o', color='red')
plt.title('Metode Elbow untuk Pemilihan Jumlah Cluster Optimal')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Inersia')
plt.grid(True)
plt.show()
# Plotting Silhouette Score
plt.figure(figsize=(10, 5))
plt.plot(range(2, 11), silhouette_avg, marker='s', color='blue')
plt.title('Silhouette Score untuk Pemilihan Jumlah Cluster Optimal')
plt.xlabel('Jumlah Cluster (k)')
plt.ylabel('Silhouette Score')
plt.grid(True)
plt.show()
```

Gambar 4. 20 Hasil Source Code Jumlah Optimal

1. Nilai *Silhoutte Coefficient*

Silhouette Coefficient merupakan metrik evaluasi pengelompokan yang digunakan untuk mengukur seberapa baik pengelompokan tersebut. *Silhoutte Coefficient* menggabungkan dua aspek yakni seberapa dekat titik-titik data dalam sebuah cluster, dan seberapa terpisah cluster tersebut dari cluster lainnya. Berdasarkan hasil perhitungan *silhoutte_score* yang diperoleh dengan menggunakan data tahun 2020-2022 Seperti pada gambar berikut:

	Number of Clusters	Inertia	Silhouette Score
0	2	434.604761	0.539527
1	3	265.780505	0.517257
2	4	191.103150	0.528109
3	5	153.068477	0.485191
4	6	120.167913	0.467346
5	7	104.883884	0.442292
6	8	94.142310	0.425721
7	9	85.001891	0.416871
8	10	75.165601	0.399429

Gambar 4. 21 Hasil dari *Silhoutte_score* Data 2020

	Number of Clusters	Inertia	Silhouette Score
0	2	393.825266	0.591120
1	3	211.520759	0.601760
2	4	154.084537	0.611351
3	5	119.573303	0.547312
4	6	98.575208	0.537059
5	7	79.627131	0.488045
6	8	66.475406	0.504037
7	9	58.663166	0.501318
8	10	50.576063	0.512165

Gambar 4. 22 Hasil dari Silhoutte_score Data 2021

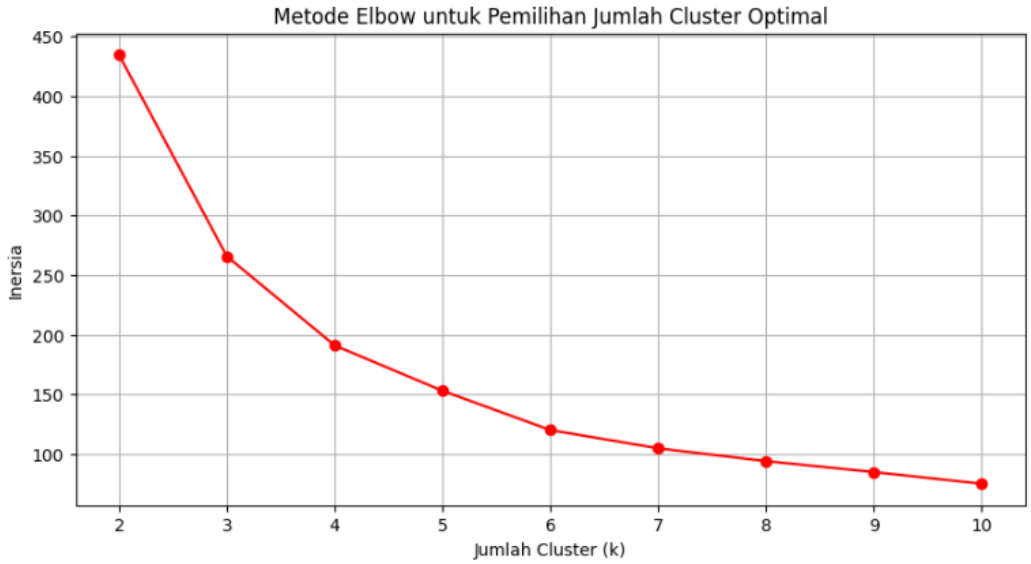
	Number of Clusters	Inertia	Silhouette Score
0	2	467.482269	0.534310
1	3	281.809170	0.501565
2	4	232.510772	0.492365
3	5	188.270444	0.449426
4	6	154.789560	0.425973
5	7	130.121759	0.405645
6	8	119.293273	0.386039
7	9	110.330957	0.372541
8	10	101.962614	0.371103

Gambar 4. 23 Hasil dari Silhoutte_score Data 2022

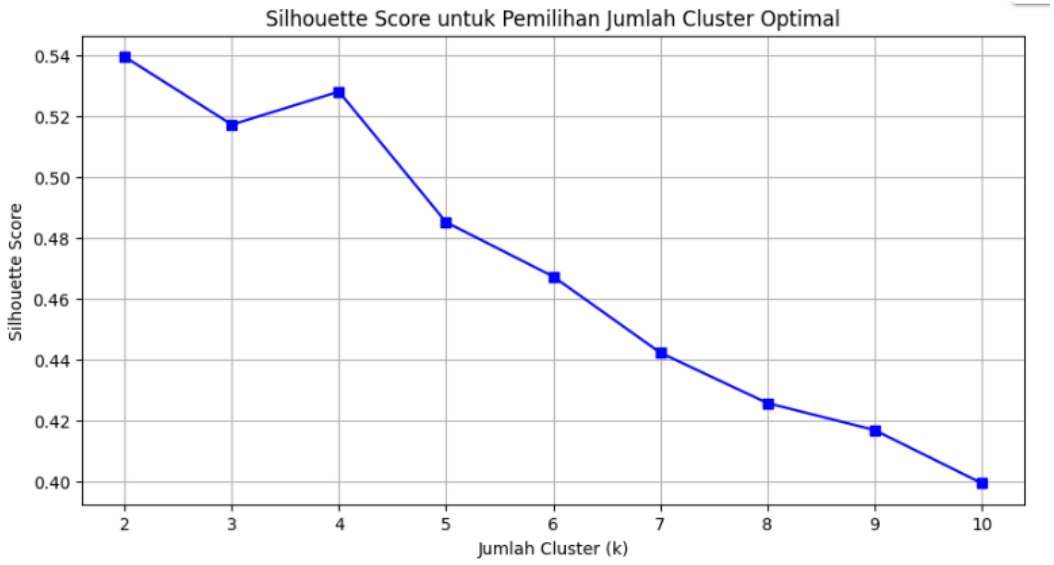
2. Metode Silhouette Coefficient dan Metode Elbow

Penentuan jumlah cluster optimal menggunakan metode Elbow dari hasil perhitungan nilai Silhoutte Coefficient. Penggabungan metode tersebut bertujuan untuk memperoleh cluster yang baik dan berkualitas. Berikut merupakan tahapan yang dilakukan untuk menentukan nilai k yang efektif dengan menggunakan silhoute coeficient dan metode Elbow.

Berikut merupakan hasil visualisasi dari metode silhouette coefficient dan metode elbow pada tahun 2020 seperti gambar berikut:

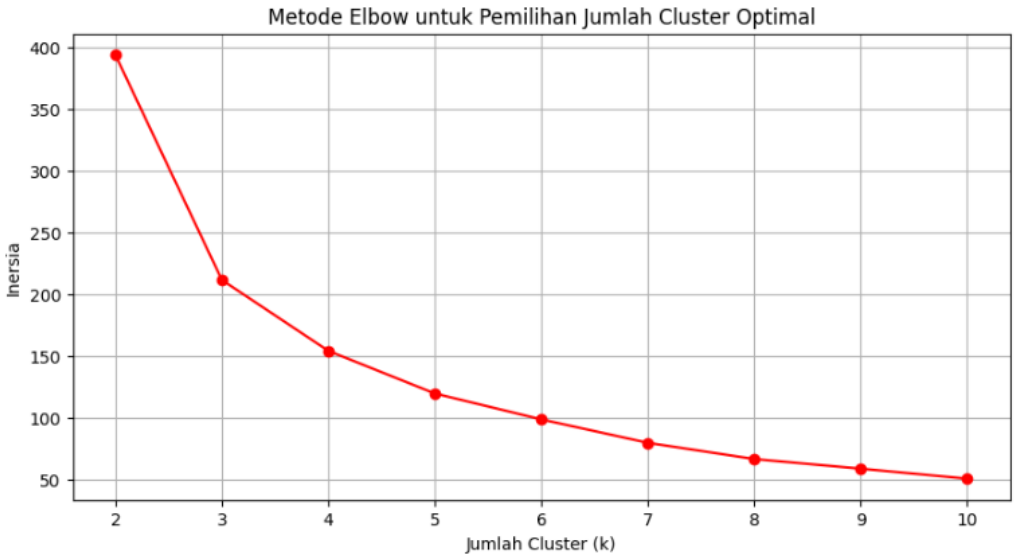


Gambar 4. 24 Hasil Visualisasi Data Jumlah Optimal 2020

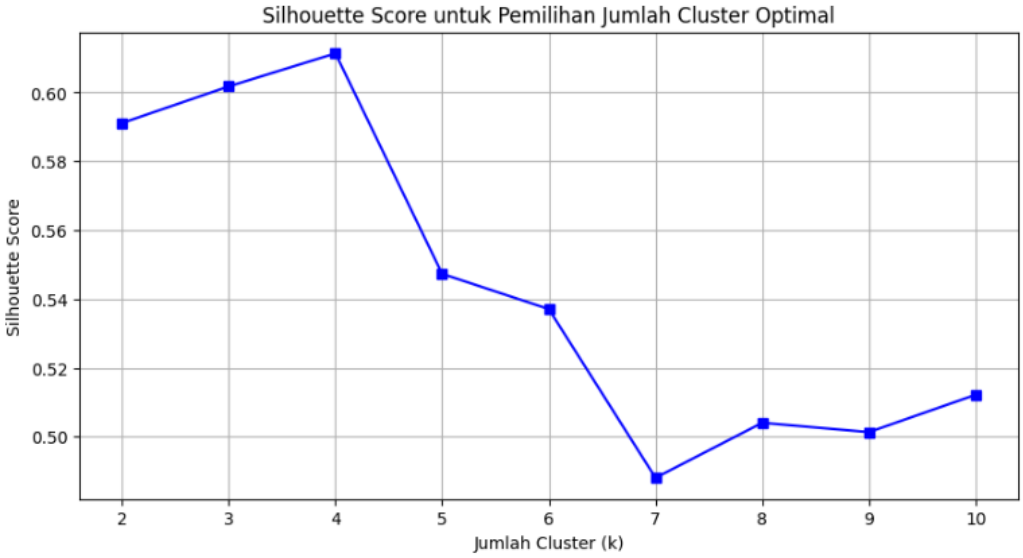


Gambar 4. 25 Hasil Visualisasi Data Jumlah Optimal 2020

Berikut merupakan hasil visualisasi dari metode silhouette coefficient dan metode elbow pada tahun 2021 seperti gambar berikut:

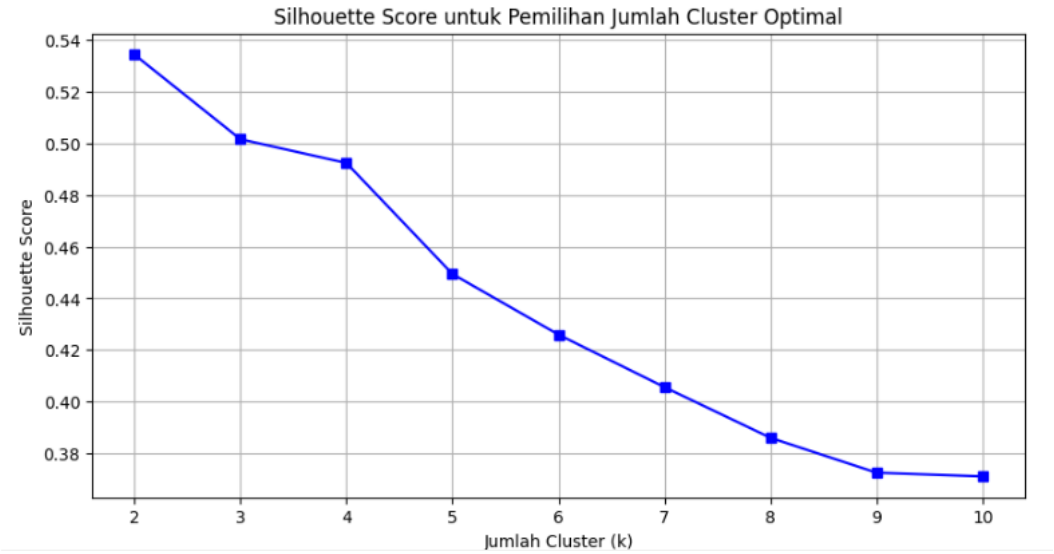


Gambar 4. 26 Hasil Visualisasi Data Jumlah Optimal 2021

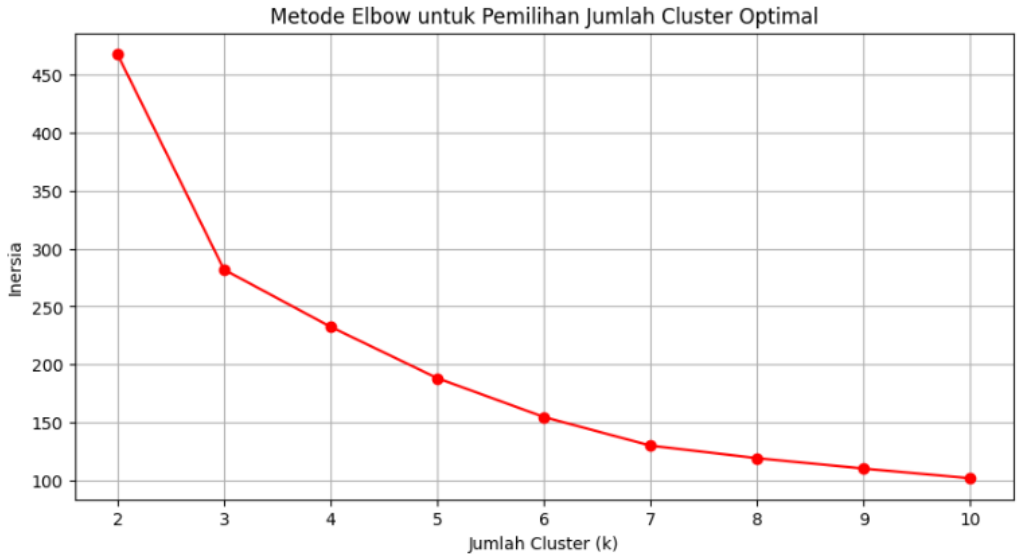


Gambar 4. 27 Hasil Visualisasi Data Jumlah Optimal 2021

Berikut merupakan hasil visualisasi dari metode silhouette coefficient dan metode elbow pada tahun 2022 seperti gambar berikut:



Gambar 4. 28 Hasil Visualisasi Data Jumlah Optimal 2022



Gambar 4. 29 Hasil Visualisasi Data Jumlah Optimal 2022

Jumlah cluster optimal yang dipilih adalah 4, berdasarkan analisis posisi siku dalam grafik elbow yang menunjukkan penurunan yang signifikan dalam nilai inertia seiring dengan peningkatan jumlah cluster. Dalam analisis klaster, inertia adalah ukuran yang mengindikasikan

seberapa jauh titik data tersebar di dalam sebuah cluster. Inertia cenderung menurun ketika jumlah cluster bertambah, penurunan tersebut melambat saat jumlah cluster mencapai jumlah optimal. Titik di mana penurunan inertia menjadi kurang signifikan menandakan jumlah cluster yang optimal. (saffar, 2021)

4.2.7 **Menjalankan Kmeans Clustering**

Berikut pada baris pertama di gambar 4.30 Mengimpor kelas KMeans dari modul cluster dalam pustaka scikit-learn. Baris kedua Menentukan jumlah klaster yang dianggap optimal (dapat berdasarkan metode Elbow atau evaluasi Silhouette). Baris ketiga Membuat objek KMeans dengan jumlah klaster yang optimal dan nilai random state untuk reproduksibilitas. Baris ke empat Melatih model KMeans pada data yang telah dinormalisasi (df_normalized). Baris kelima Memprediksi klaster untuk setiap sampel dalam data dengan menambahkan 1 karena label klaster dimulai dari 0. Baris keenam Menambahkan kolom 'Cluster' yang berisi label klaster ke dalam dataframe 'df_normalized' dan 'data'. Baris ketujuh Mencetak informasi tentang model KMeans yang telah dilatih. Baris kedelapan Menyimpan dataframe 'data' yang telah ditambahkan kolom klasternya ke dalam file Excel dengan nama 'data_cluster.xlsx'.

```
from sklearn.cluster import KMeans
optimal_k = 4
kmeans = KMeans(n_clusters=optimal_k, random_state=0 , init = 'k-means++',max_iter=100)
kmeans.fit(df_normalized)
cluster_labels = kmeans.labels_ + 1
data['Cluster'] = cluster_labels
print("Model KMeans yang telah dilatih:", kmeans)
data.to_excel('data_cluster.xlsx')
data
```

Gambar 4. 30 Sourcing Code Clustering

Berdasarkan hasil penentuan jumlah cluster optimal menggunakan metode Elbow dilakukan clusterisasi dengan nilai k = 4 untuk data tahun 2022. Hasil clustering yang telah terbentuk kemudian disimpan dengan menambahkan kolom baru pada dataframe yang sudah dibuat sebelumnya. Hal ini bertujuan untuk memudahkan visualisasi data dan analisis perubahan cluster. Pada DataFrame tersebut akan diurutkan berdasarkan kolom cluster yang terbentuk. Berikut adalah hasil clustering data tahun 2020-2022.

Berikut merupakan hasil clustering data pada tahun 2020 seperti pada gambar berikut:

	no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah	Cluster
0	1	Cerai Gugat	30	31	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Sambeng, Kabupat...	3
1	2	Cerai Gugat	27	33	Ekonomi	Kantor Urusan Agama Kecamatan Bluluk, Kabupate...	1
2	3	Cerai Gugat	38	51	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Paciran, Kabupat...	3
3	4	Cerai Gugat	26	27	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama (KUA) Kecamatan Solokuro K...	3
4	5	Cerai Gugat	26	36	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Brondong, Kabupa...	3
...	...	...	...	...	...	...	...
2480	2481	Cerai Gugat	63	50	Ekonomi	KUA KecamatanSekaran, Kabupaten Lamongan	1
2481	2482	Cerai Gugat	52	53	Zina	KUA Kecamatan Modo, Kabupaten Lamongan	3
2482	2483	Cerai Gugat	27	44	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Modo, Kabupaten Lamongan	3
2483	2484	Cerai Gugat	41	38	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Mantup, Kabupaten Lamongan	3
2484	2485	Cerai Gugat	39	33	Ekonomi	KUA Kecamatan Ngimbang, Kabupaten Lamongan	1

2485 rows × 7 columns

Gambar 4. 31 Hasil Clustering Data Tahun 2020

Berikut merupakan hasil clustering data pada tahun 2020 seperti pada gambar berikut:

	no	jenis_cerai	umur_p	umur_t	faktor_perceraian	kua_tempat_nikah	Cluster
0	1	Cerai Talak	26	25	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Kembang bahu Kab...	2
1	2	Cerai Talak	48	40	Ekonomi	KUA Kecamatan Nanga Pinoh, Kabupaten Sintang	4
2	3	Cerai Talak	50	41	Ekonomi	KUA Kecamatan Modo, Kabupaten Lamongan	4
3	4	Cerai Talak	42	38	Peselisihan dan Pertengkaran Terus Menerus	Kantor Urusan Agama Kecamatan Duduksampeyan, K...	2
4	5	Cerai Talak	27	23	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Kedungpring, Kabupaten Lamongan	2
...	...	...	...	...	...	...	...
2678	2679	Cerai Gugat	35	40	Meninggalkan Salah Satu Pihak	KUA Kecamatan Karangbinangunj, Kabupaten Lamongan	3
2679	2680	Cerai Gugat	35	47	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Sukorame, Kabupaten Lamongan	3
2680	2681	Cerai Gugat	26	29	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Bluluk, Kabupaten Lamongan	3
2681	2682	Cerai Gugat	31	32	Peselisihan dan Pertengkaran Terus Menerus	KUA Kecamatan Mantup, Kabupaten Lamongan	3
2682	2683	Cerai Gugat	45	42	Ekonomi	KUA Kecamatan Modo, Kabupaten Lamongan	1

2683 rows × 7 columns

Gambar 4. 32 Hasil Clustering Data Tahun 2021

Berikut merupakan hasil clustering data pada tahun 2022 seperti pada gambar berikut:

	no	jenis_cerai	angka jenis	umur_p	umur_t	faktor_perceraian	angka_faktor	kua_tempat_nikah	angka_kua_tempat_nikah	Cluster
0	710	Cerai Gugat	1	56	78	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Gumukmas, Kabupaten Jember	29	1
1	1738	Cerai Talak	2	78	78	Meninggalkan Salah Satu Pihak	3	KUA Kecamatan Maduran, Kabupaten Lamongan	13	1
2	785	Cerai Gugat	1	73	77	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Laren Kabupaten Lamongan	12	1
3	1956	Cerai Gugat	1	49	74	Kekerasan Dalam Rumah Tangga	4	KUA Kecamatan Deket Kabupaten Lamongan	4	1
4	982	Cerai Gugat	1	63	71	Ekonomi	2	KUA Kecamatan Kedungpring, Kabupaten Lamongan	9	1
...	...	...	...	...	...	...	...	...	...	...
2599	2215	Cerai Talak	2	27	18	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Modo Kabupaten Lamongan	15	3
2600	2299	Cerai Talak	2	38	17	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Krembangan, Kota Surabaya	29	3
2601	2348	Cerai Gugat	1	17	17	Ekonomi	2	KUA Kecamatan Mantup, Kabupaten Lamongan	14	4
2602	2556	Cerai Talak	2	25	17	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Sambeng	19	3
2603	658	Cerai Talak	2	20	16	Peselisihan dan Pertengkaran Terus Menerus	1	KUA Kecamatan Pucuk, Kabupaten Lamongan	18	3

Gambar 4. 33 Hasil Clustering Data Tahun 2022

4.2 Evaluasi dan Hasil Analisis Cluster

4.2.1 Evaluasi Cluster

Evaluasi klaster bertujuan untuk menilai seberapa baik klaster yang terbentuk. Perhitungan kualitas klaster dalam penelitian ini menggunakan Silhouette Coeffisien. Silhouette Coeffisien memiliki rentang nilai dari -1 hingga 1, di mana nilai +1 menandakan pengelompokan yang sangat baik, sementara nilai -1 menandakan pengelompokan yang buruk. Nilai 0 menunjukkan titik data yang berada tepat di antara dua klaster.

Berdasarkan hasil nilai silhoutte score yang diperoleh data tahun 2020 yang bernilai dengan rentan 0,39 – 0,53 tahun 2021 bernilai dengan rentan 0,50 – 0,61, dan pada tahun 2022 bernilai dengan rentan 0,37 – 0,53 seperti pada gambar 4. 18, 4.19 dan 4.20 maka clustering yang terbentuk memiliki interpretasi yang cukup baik karena mendekati angka 1.

4.2.1 Hasil Analisis Cluster

Berdasarkan hasil clustering yang dihasilkan menggunakan program Python, berikut adalah hasil clustering yang diperoleh dari data tahun 2020 - 2022 dengan jumlah klaster (k) sebanyak 4, seperti yang terdapat pada tabel.

Tabel 4. 4 Hasil Clustering Data 2020

Hasil Klaster Data Tahun 2020				
Cluster_1				
Jumlah Anggota 878				
Kua_Tempat_Nikah	Jumlah	Presentase	Faktor_Penyebab_Terbanyak	Jumlah_Terbanyak
KUA Kecamatan Babat	56	6,37%	Ekonomi	51

KUA Kecamatan kedungpring	37	4,21%	Ekonomi	34
KUA Kecamatan Paciran	34	3,87%	Ekonomi	34
KUA Kecamatan Lamongan	32	3,64%	Ekonomi	31
KUA Kecamatan sukodadi	30	3,41%	Ekonomi	25
Cluster_2				
Jumlah Anggota 320				
KUA Kecamatan Kedungpring	16	0,05%	Ekonomi	16
KUA Kecamatan Ngimbang	15	4,68%	Ekonomi	15
KUA Kecamatan Kembangbahu	13	4,06%	Ekonomi	12
KUA Kecamatan Sugio	12	3,75%	Ekonomi	12
KUA Kecamatan Sambeng	11	3,43%	Ekonomi	10
Cluster_3				
Jumlah Anggota 856				
KUA Kecamatan Babat	46	5,37%	Peselisihan dan Pertengkaran terus Menerus	30
KUA Kecamatan Brondong	43	5,02%	Peselisihan dan Pertengkaran terus Menerus	35
KUA Kecamatan Paciran	43	5,02%	Peselisihan dan Pertengkaran terus Menerus	30
KUA Kecamatan Lamongan	33	3,85%	Peselisihan dan Pertengkaran terus Menerus	30
KUA Kecamatan Deket	29	3,38%	Peselisihan dan Pertengkaran terus Menerus	25
Cluster_4				
Jumlah Anggota 431				

KUA Kecamatan Paciran	19	4,40%	Peselisihan dan pertengakaran terus menerus	16
KUA Kecamatan Sugio	18	4,17%	Peselisihan dan pertengakaran terus menerus	16
KUA Kecamatan Brondong	15	3,48%	Peselisihan dan pertengakaran terus menerus	14
KUA Kecamatan Sekaran	14	3,24%	Peselisihan dan pertengakaran terus menerus	12
KUA Kecamatan Lamongan	14	3,24%	Peselisihan dan pertengakaran terus menerus	10

Tabel 4. 5 Hasil Clustering Data 2021

Hasil Klaster Data Tahun 2021				
Cluster_1				
Jumlah Anggota 800				
Kua_Tempat_Nikah	Jumlah	Presentase	Faktor_Penyebab_Terbanyak	Jumlah_Terbanyak
KUA Kecamatan Paciran	45	5,62%	Ekonomi	33
KUA Kecamatan Babat	39	4,87%	Ekonomi	34
KUA Kecamatan Brondong	34	4,25%	Ekonomi	27
KUA Kecamatan Kedungpring	33	4,12%	Ekonomi	28
KUA Kecamatan Sugio	33	4,12%	Ekonomi	30
Cluster_2				
Jumlah Anggota 546				
KUA Kecamatan Kedungpring	23	4,21%	Peselisihan dan Pertengkaran Terus Menenrus	21
KUA Kecamatan Sukodadi	22	4,02%	Peselisihan dan Pertengkaran Terus Menenrus	18
KUA Kecamatan Paciran	20	3,66%	Peselisihan dan Pertengkaran Terus Menenrus	18

KUA Kecamatan Mantup	19	3,47%	Peselisihan dan Pertengkaran Terus Menenrus	17
KUA Kecamatan Babat	18	3,29%	Peselisihan dan Pertengkaran Terus Menenrus	16
Cluster_3				
Jumlah Anggota 1156				
KUA Kecamatan Kedungpring	49	4,23%	Peselisihan dan Pertengkaran terus Menerus	45
KUA Kecamatan Babat	49	4,23%	Peselisihan dan Pertengkaran terus Menerus	46
KUA Kecamatan Mantup	49	4,23%	Peselisihan dan Pertengkaran terus Menerus	42
KUA Kecamatan Sambeng	44	3,80%	Peselisihan dan Pertengkaran terus Menerus	42
KUA Kecamatan Paciran	44	3,80%	Peselisihan dan Pertengkaran terus Menerus	41
Cluster_4				
Jumlah Anggota 181				
KUA Kecamatan Brondong	12	6,62%	Ekonomi	12
KUA Kecamatan Sukodadi	11	6,07%	Ekonomi	11
KUA Kecamatan Kembangbahu	9	4,97%	Ekonomi	8
KUA Kecamatan Sambeng	7	3,88%	Ekonomi	7
KUA Kecamatan Paciran	6	3,31%	Ekonomi	6

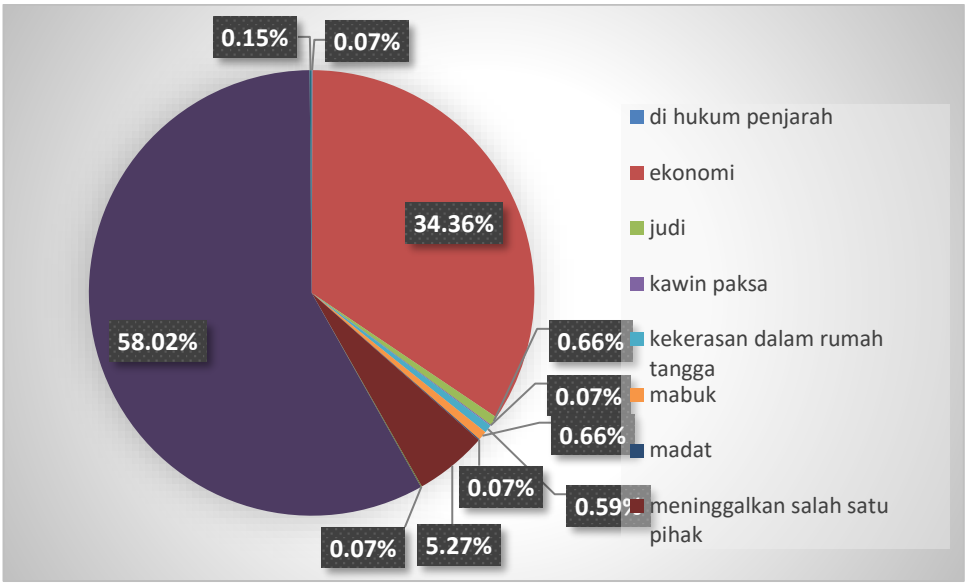
Tabel 4. 6 Hasil Clustering Data 2022

Hasil Klaster Data Tahun 2022				
Cluster_1				
Jumlah Anggota 1019				
Kua_Tempat_Nikah	Jumlah	Presentase	Faktor_Penyebab_Terbanyak	Jumlah_Terbanyak
KUA Kecamatan Babat	54	5,29%	Peselisihan dan Pertengkaran Terus Menerus	40
KUA Kecamata Brondong	49	4,80%	Peselisihan dan Pertengkaran Terus Menerus	41
KUA Kecamatan Paciran	48	4,71%	Peselisihan dan Pertengkaran Terus Menerus	40
KUA Kecamatan Lamongan	37	3,63%	Peselisihan dan Pertengkaran Terus Menerus	32
KUA Kecamatan Ngimbang	34	3,33%	Peselisihan dan Pertengkaran Terus Menerus	28
Cluster_2				
Jumlah Anggota 559				
KUA Kecamatan Babat	28	5,00%	Peselisihan dan Pertengkaran Terus Menenrus	26
KUA Kecamatan Brondong	25	4,47%	Peselisihan dan Pertengkaran Terus Menenrus	17
KUA Kecamatan Sekaran	21	3,75%	Peselisihan dan Pertengkaran Terus Menenrus	21
KUA Kecamatan Sugio	21	3,75%	Peselisihan dan Pertengkaran Terus Menenrus	16
KUA Kecamatan Paciran	19	3,39%	Peselisihan dan Pertengkaran Terus Menenrus	15
Cluster_3				
Jumlah Anggota 881				
KUA Kecamatan Babat	39	4,42%	Ekonomi	34
KUA Kecamatan Sambeng	36	4,09%	Ekonomi	31
KUA Kecamatan Kembangbahu	34	3,86%	Ekonomi	30
KUA Kecamatan Sugio	32	3,63%	Ekonomi	30
KUA Kecamatan Paciran	30	3,40%	Ekonomi	22
Cluster_4				
Jumlah Anggota 149				
KUA Kecamatan Brondong	14	9,39%	Ekonomi	14

KUA Kecamatan Babat	8	5,36%	Ekonomi	8
KUA Kecamatan Paciran	8	5,36%	Ekonomi	8
KUA Kecamatan Sugio	6	4,02%	Ekonomi	6
KUA Kecamatan Ngimbang	6	4,02%	Ekonomi	6

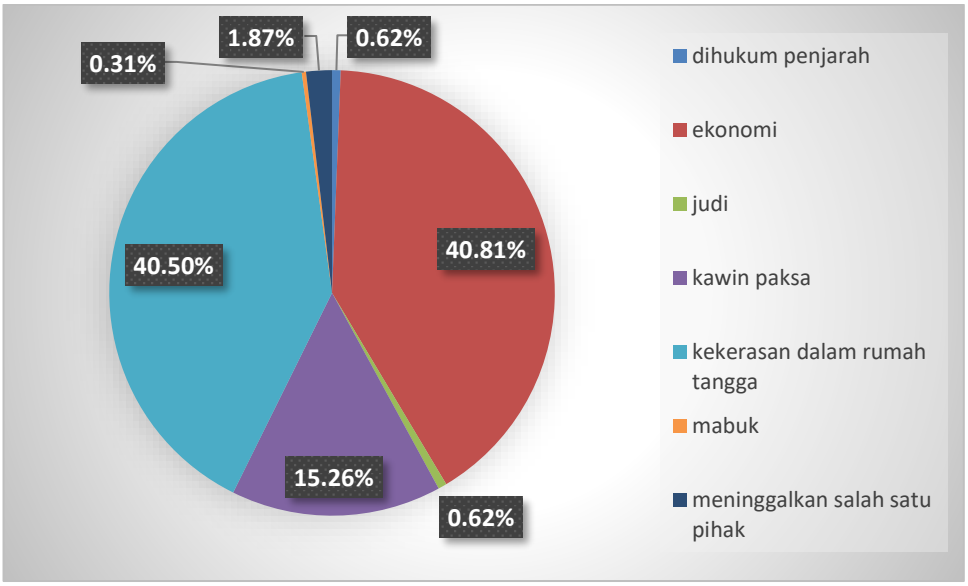
Berdasarkan hasil Clustering dengan data tahun 2020, terbentuk empat cluster dengan karakteristik sebagai berikut:

Cluster_1 : Data Cluster 1 adalah 878 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 21,5% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempah nikah. pada cluster 1 kecamatan dengan kasus tertinggi yang ke satu yaitu babat mencapai 6,37% total kasus, yang ke dua yaitu kedunpring sebesar 4,21% total kasus, yang ke tiga paciran dengan 3,87% total kasus, yang ke empat lamongan dengan nilai sebesar 3,64% total kasus dan yang terakhir yaitu sukodadi 3,41% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart



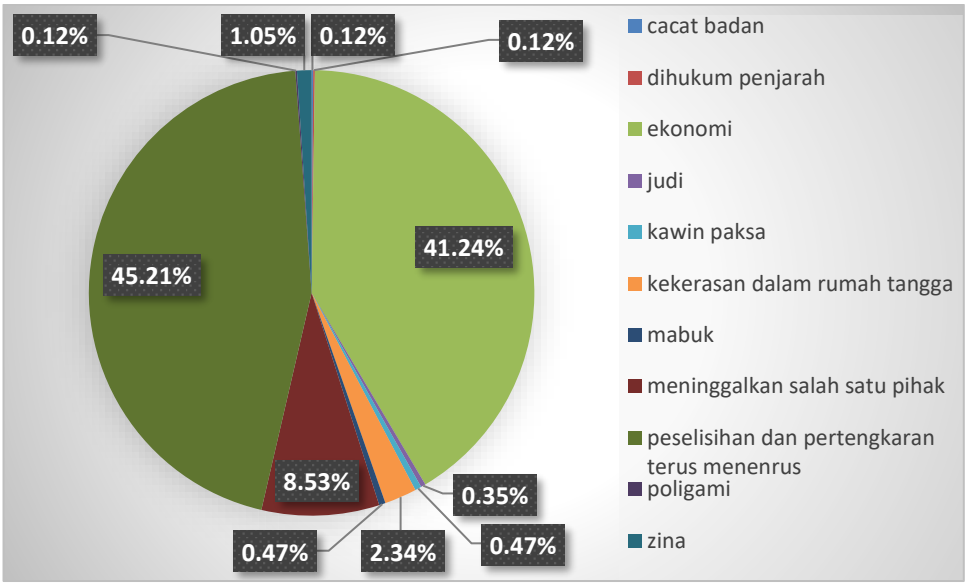
Gambar 4. 34 hasil chart claster 1

Cluster_2 : Data Cluster 2 adalah 320 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 15,97% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 2 kecamatan dengan kasus terendah yang ke satu yaitu kedungpring mencapai 0,05% total kasus, yang ke dua yaitu ngimbang sebesar 4,68% total kasus, yang ke tiga kembangbahu dengan 4,06% total kasus, yang ke empat sugio dengan nilai sebesar 3,75% total kasus dan yang terakhir yaitu sambeng 3,43% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



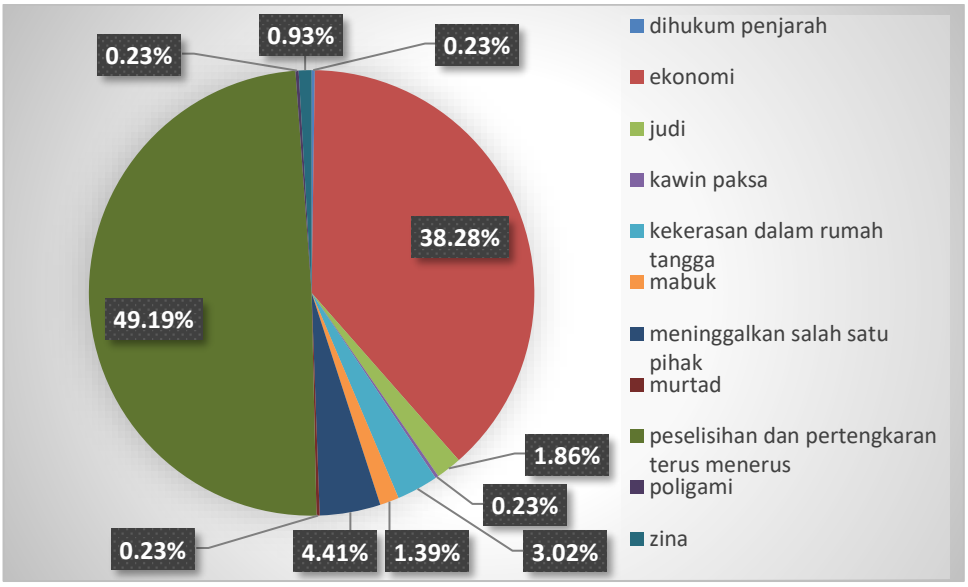
Gambar 4. 35 hasil chart claster 2

Cluster_3 Data Cluster 3 adalah 856 anggota yang memiliki faktor penyebab utama yaitu masalah peselisihan dan pertengkaran terus menerus dengan nilai sebesar 22,64% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 3 kecamatan dengan kasus sedang yang ke satu yaitu babat mencapai 5,37%% total kasus, yang ke dua yaitu brondong sebesar 5,02% total kasus, yang ke tiga paciran dengan 5,02% total kasus, yang ke empat lamongan dengan nilai sebesar 3,85% total kasus dan yang terakhir yaitu deket 3,38% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



Gambar 4. 36 hasil chart claster 3

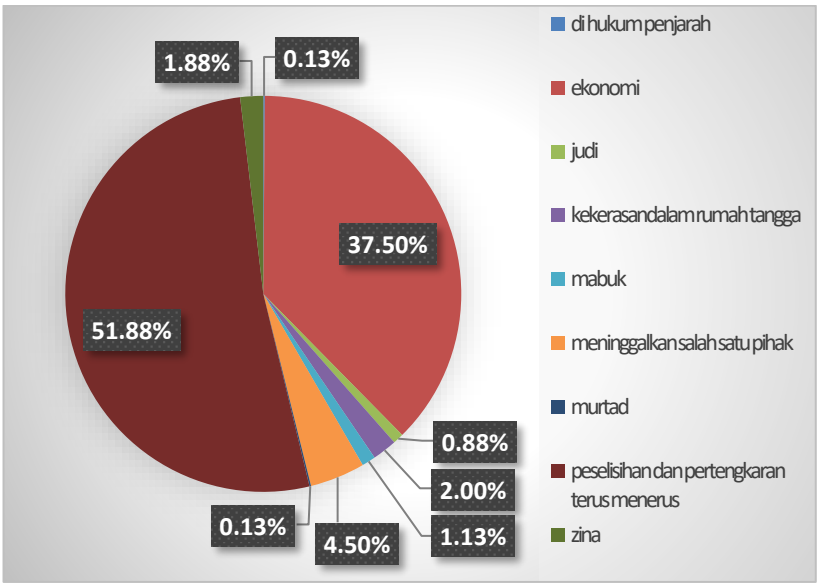
Cluster_4. Data Cluster 4 adalah 431 anggota yang memiliki faktor penyebab utama yaitu masalah peselisihan dan pertengkaran terus menerus dengan nilai sebesar 18,53% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 4 kecamatan dengan kasus tinggi yang ke satu yaitu paciran 4,40% total kasus, yang ke dua yaitu sugio sebesar 4,17% total kasus, yang ke tiga brondong dengan 3,48% total kasus, yang ke empat sekaran dengan nilai sebesar 3,24% total kasus dan yang terakhir yaitu lamongan 3,24% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



Gambar 4. 37 hasil chart claster 4

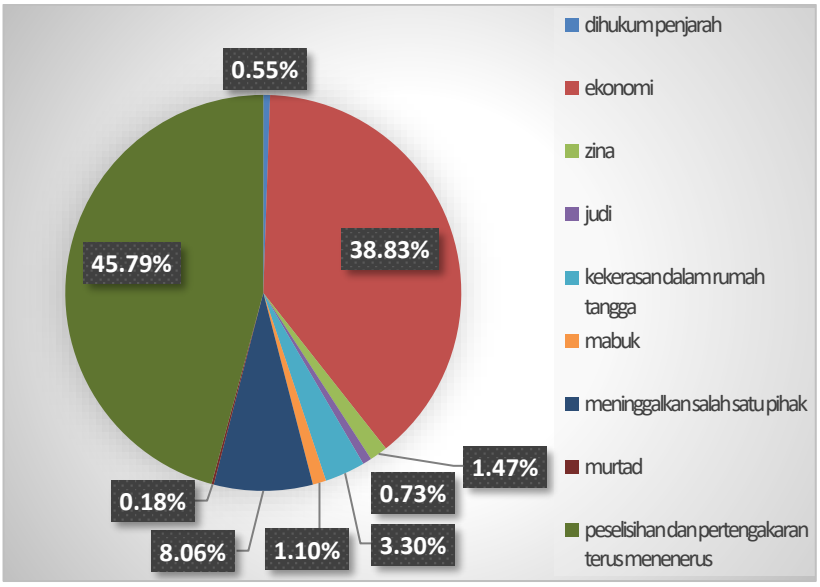
Berdasarkan hasil Clustering dengan data tahun 2021, terbentuk empat cluster dengan karakteristik sebagai berikut:

Cluster_1 Data Cluster 1 adalah 800 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 22,98% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 1 kecamatan dengan kasus tertinggi yang ke satu yaitu paciran mencapai 5,62% total kasus, yang ke dua yaitu babat sebesar 4,87% total kasus, yang ke tiga brondong dengan 4,25% total kasus, yang ke empat kedungpring dengan nilai sebesar 4,12% total kasus dan yang terakhir yaitu sugio 4,12% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



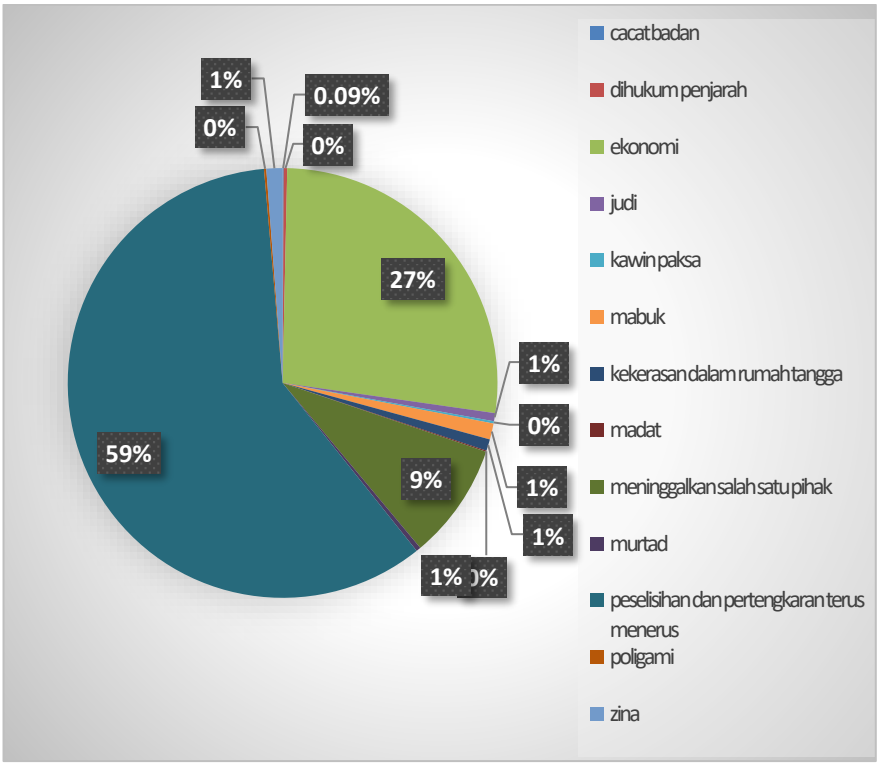
Gambar 4. 38 hasil chart claster 1

Cluster_2 Data Cluster 2 adalah 546 anggota yang memiliki faktor penyebab utama yaitu masalah peselisihan dan pertengkaran terus menerus dengan nilai sebesar 18,59% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 2 kecamatan dengan kasus tertinggi yang ke satu yaitu kedungpring mencapai 4,21% total kasus, yang ke dua yaitu sukodadi sebesar 4,02% total kasus, yang ke tiga paciran dengan 3,66% total kasus, yang ke empat mantup dengan nilai sebesar 3,47% total kasus dan yang terakhir yaitu babat 3,29% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



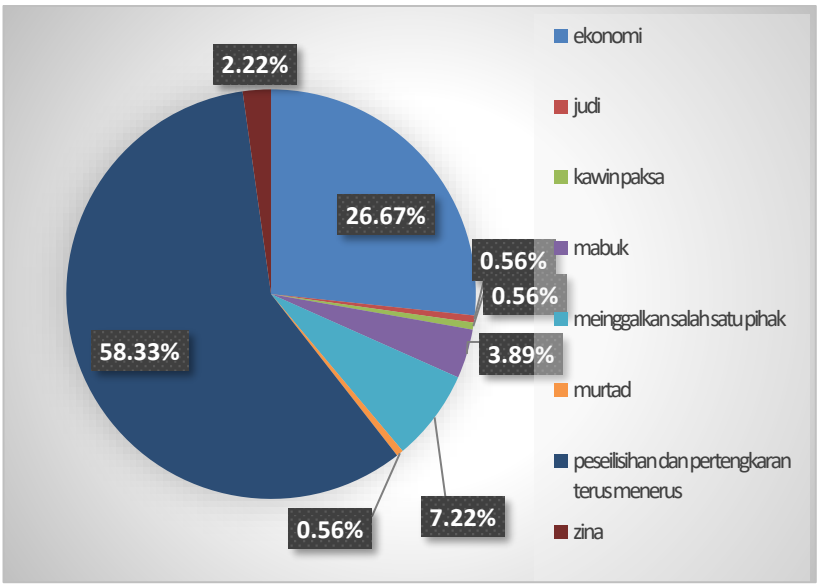
Gambar 4. 39 hasil chart claster 2

Cluster_3 Data Cluster 3 adalah 1156 anggota yang memiliki faktor penyebab utama yaitu masalah peselisihan dan pertengkaran terus menerus dengan nilai sebesar 20,29% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 3 kecamatan dengan kasus sedang yang ke satu yaitu kedungpring mencapai 4,23% total kasus, yang ke dua yaitu babat sebesar 4,23% total kasus, yang ke tiga mantup dengan 4,23% total kasus, yang ke empat sambeng dengan nilai sebesar 3,80% total kasus dan yang terakhir yaitu paciran 3,80% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



Gambar 4. 40 hasil chart claster 3

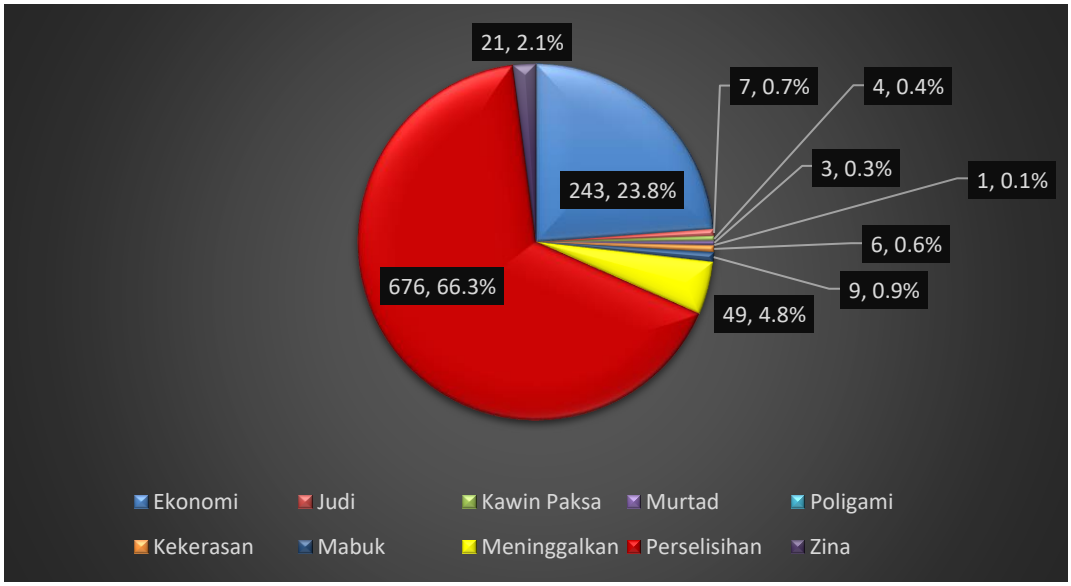
Cluster_4 Data Cluster 4 adalah 181 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 24,85% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 4 kecamatan dengan kasus tertinggi yang ke satu yaitu brondong mencapai 6,62% total kasus, yang ke dua yaitu sukodadi sebesar 6,07% total kasus, yang ke tiga kembangbahu dengan 4,97% total kasus, yang ke empat sambeng dengan nilai sebesar 3,88% total kasus dan yang terakhir yaitu paciran 3,31% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



Gambar 4. 41 hasil chart claster 4

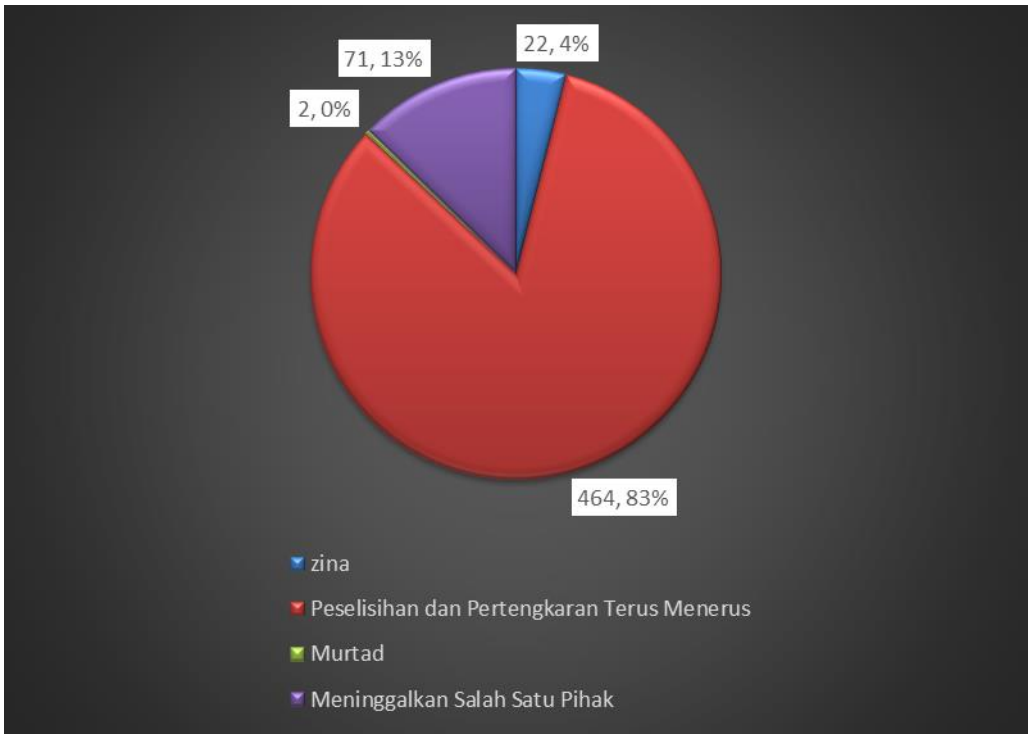
Berdasarkan hasil Clustering dengan data tahun 2022, terbentuk empat cluster dengan karakteristik sebagai berikut:

Cluster_1 Data Cluster 1 adalah 1019 anggota yang memiliki faktor penyebab utama yaitu masalah peselisihan dan pertengkaran terus menerus dengan nilai sebesar 21,76% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 1 kecamatan dengan kasus tertinggi yang ke satu yaitu babat mencapai 5,29% total kasus, yang ke dua yaitu brondong sebesar 4,80% total kasus, yang ke tiga paciran dengan 4,71% total kasus, yang ke empat lamongan dengan nilai sebesar 3,63% total kasus dan yang terakhir yaitu ngimbang 3,33% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



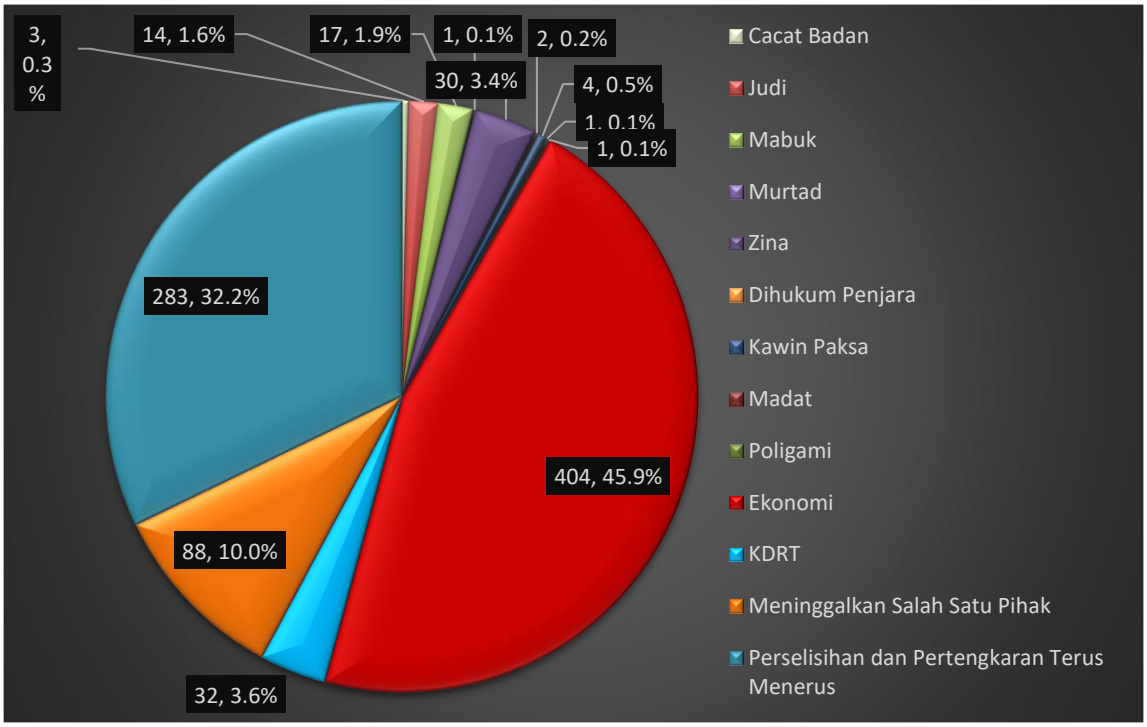
Gambar 4. 42 hasil chart claster 1

Cluster_2 Data Cluster 2 adalah 559 anggota yang memiliki faktor penyebab utama yaitu masalah peselisiha dan pertengakaran terus menrus dengan nilai sebesar 20,36% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 2 kecamatan dengan kasus tertinggi yang ke satu yaitu babat mencapai 5,00% total kasus, yang ke dua yaitu brondong sebesar 4,47% total kasus, yang ke tiga sekaran dengan 3,75% total kasus, yang ke empat sugio dengan nilai sebesar 3,75% total kasus dan yang terakhir yaitu paciran 3,39% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



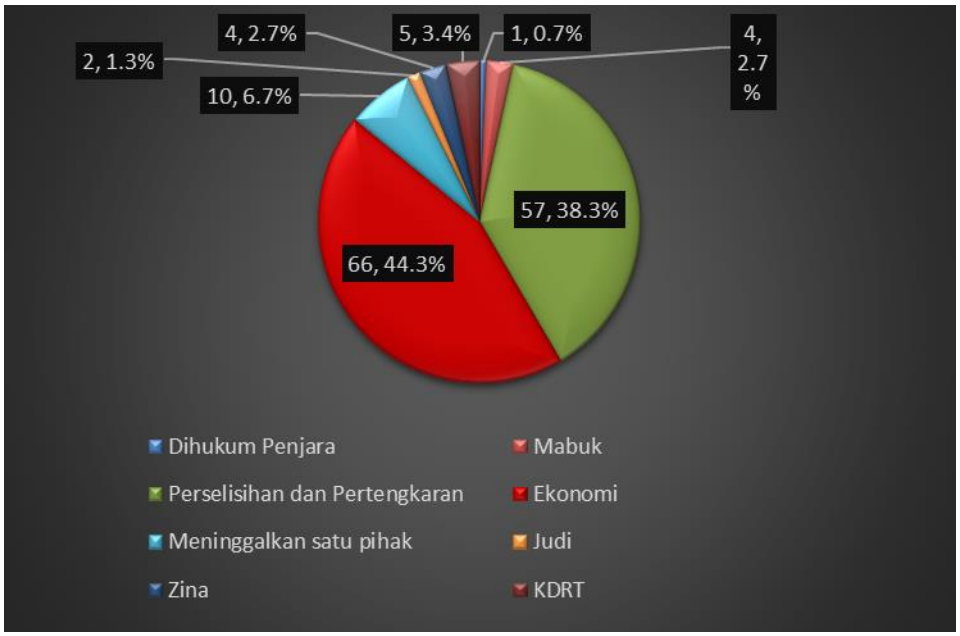
Gambar 4. 43 hasil chart claster 2

Cluster_3 Data Cluster 3 adalah 881 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 19,4% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 3 kecamatan dengan kasus tertinggi yang ke satu yaitu babat mencapai 4,42% total kasus, yang ke dua yaitu sambeng sebesar 4,09% total kasus, yang ke tiga kembangbahu dengan 3,86% total kasus, yang ke empat sugio dengan nilai sebesar 3,63% total kasus dan yang terakhir yaitu paciran 3,40% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart



Gambar 4. 44 hasil chart claster 3

Cluster_4 Data Cluster 4 adalah 149 anggota yang memiliki faktor penyebab utama yaitu masalah ekonomi dengan nilai sebesar 19,15% keseluruhan total kasus. ada 5 jenis variabel atau atribut yang digunakan untuk menentukan setiap cluster yaitu Jenis cerai, umur penggugat, umur tergugat, faktor penyebab dan KUA tempat nikah. pada cluster 4 kecamatan dengan kasus tertinggi yang ke satu yaitu brondong mencapai 9,39% total kasus, yang ke dua yaitu babat sebesar 5,36% total kasus, yang ke tiga paciran dengan 5,36% total kasus, yang ke empat sugio dengan nilai sebesar 4,02% total kasus dan yang terakhir yaitu ngimbang 4,02% total kasus. Berikut merupakan hasil presentase dari faktor penyebab seperti pada gambar chart.



Gambar 4. 45 hasil chart claster 4

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan latar belakang yang telah diuraikan sebelumnya, penelitian ini bertujuan untuk melakukan klasterisasi data perceraian di wilayah Kabupaten Lamongan berdasarkan atribut jenis cerai, umur penggugat, umur tergugat, faktor penyebab, dan KUA tempat nikah/kecamatan. Bertujuan untuk Melakukan analisis data perceraian di pengadilan agama lamongan dengan melakukan klasterisasi K-means guna mengelompokan daerah yang memiliki kasus perceraian tertinggi berdasarkan karakteristik yang ditentukan

Berdasarkan pembahasan dan hasil implementasi menggunakan Collab Netbook pada klasterisasi data perceraian di wilayah Kabupaten Lamongan, dapat disimpulkan bahwa terdapat 4 klaster pada tahun 2020 - 2022. yaitu data hasil analisis perceraian pada kecematan babat, Paciran dan brondong sebagai lokasi dengan kasus perceraian tertinggi memiliki karakteristik yang telah ditentukan pada setiap kecamatan. jadi faktor utama perceraian yang terjadi pada kecamatan babat, Paciran dan brondong di sebabkan oleh ekonomi, perselisihan dan pertengkaran yang terjadi terus menerus. pada beberapa kasus perceraian yang telah terjadi. Adapun perubahan pada tahun 2022 peningkatan anggota cluster yang menyatakan peselisihan dan pertengkaran sebagai penyebab utama perceraian.

Cluster dengan anggota besar cenderung di pengaruhi oleh peselisihan dan pertengkaran. cluster dengan anggota lebih kecil menunjukkan masalah ekonomi sebagai faktor dominan

5.2 Saran

Saran penulisan terkait dengan penelitian "KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN METODE K-MEANS" untuk peneliti selanjutnya ntuk mengembangkan penelitian ini, aplikasi yang digunakan dalam analisis K-Means Clustering bisa menggunakan aplikasi yang lebih modern dan lebih canggih untuk bisa menciptakan visualisasi dari Clustering berupa gambar titik map di setiap kecamatan.

DAFTAR PUSTAKA

amelia, & ema. (2023). **Kajian Trauma Psikososial Anak Karena Perceraian Orang Tua**. jakarta: fakultad dakwah dan ilmu komunikasi universitas islam negeri syarif hidayatulllah jakarta.

binus. (2021, september 30). **Proses Data Mining KDD**. Diambil kembali dari SCHOOL OF INFORMATION SYSTEMS: ttps://sis.binus.ac.id/proses-data-mining-kdd/

buulolo, & efori. (2020). **Data Mining Untuk Perguruan Tinggi**. jl,rajawali, G. elang 6, no 3, drono, sardonoharjo, ngaklik, sleman: anggota ikpi (076/diy/2012).

firdaus, & afrizal. (2020, maret 1). **Cara kerja Kmeans**. Diambil kembali dari medium: https://medium.com/kmeans-clustering-dan-implementasinya

firdaus, & afrizal. (2020, maret 1). **Kmeans Clustering**. Diambil kembali dari medium: https://medium.com/kmeans-clustering-dan-implementasinya

Haris , k., Sarjon , D., & Sumijan. (2020). **Data Mining Menggunakan Metode K-Means Clustering** Untuk. JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST*)*, 1-119.

hidayah, & Nur Fitri. (2023, mei 22). **5 Faktor Tertinggi Penyebab Perceraian di Indonesia**. Diambil kembali dari GoodStats: https://data.goodstats.id/5-faktor-tertinggi-penyebab-perceraian-di-indonesia

insanul, k., ulya, k., & mustakin. (2019). **Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau**. jurnal ilmia rekayasa dan manajemen sistem informasi, 1-125.

jacob, & jessica. (2023, januari 13). **Manfaat Data Mining**. Diambil kembali dari taptalk.oi: https://taptalk.io/blog/manfaat-data-mining/

lamongan, p. a. (2023, oktober rabu). **Statistik Perkara.** Diambil kembali dari pa lamongan: https://pa-lamongan.go.id/pages/statistik-perkara-pengadilan-agama-lamongan

Merliana, N. E., Ernawati, & Santoso, J. (2015). **ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS**. Kajian Multi Disiplin Ilmu untuk Mewujudkan Poros Maritim dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat, 2-3.

Missaoui, R., & Abdessalem, T. (2017). **Trends in Social Network Analysis**. Switzerland: Springer.

Ni Putu , e. M., Ernawati, & Alb, j. s. (2015). **ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-MEANS**. PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU & CALL FOR PAPERS UNISBANK (SENDI_U), 3-4.

Nugroho, & Andy. (2020, February 23). **Mengenal Konsep Dasar, Tujuan, dan Teknik Data Mining**. Diambil kembali dari qwords: https://qwords.com/blog/data-mining-adalah/

nur , h., & dkk. (2022). **Analisis Metode Klastering Pada Kasus Penyebab Perceraian**. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 1.

pawestri, & hillary sekar. (2023, september 7). **8 Efek Perceraian Pada Kesehatan Fisik dan Mental**. Diambil kembali dari hellosehat.com: https://hellosehat.com/mental/hubungan-harmonis/efek-perceraian-terhadap-kesehatan/

REPUBLIK INDONESIA. (1989). **UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 7 TAHUN 1989 TENTANG PERADILAN AGAMA**. UNDANG-UNDANG REPUBLIK INDONESIA, 1-45.

Riri Narasati1, S. E. (2023). **PENERAPAN METODE TWO-STEP CLUSTER UNTUK PENGELOMPOKAN**. Jurnal Ilmiah Manajemen Informatika dan Komputer, 2.

Siti Humairoh Sa'dah1, M. S. (2019). **Analisa Metode K-Means dalam Mengelompokkan Perkara Perceraian Berdasarkan Daerah Simalungun**. Prosiding Seminar Nasional Riset Information, 1.

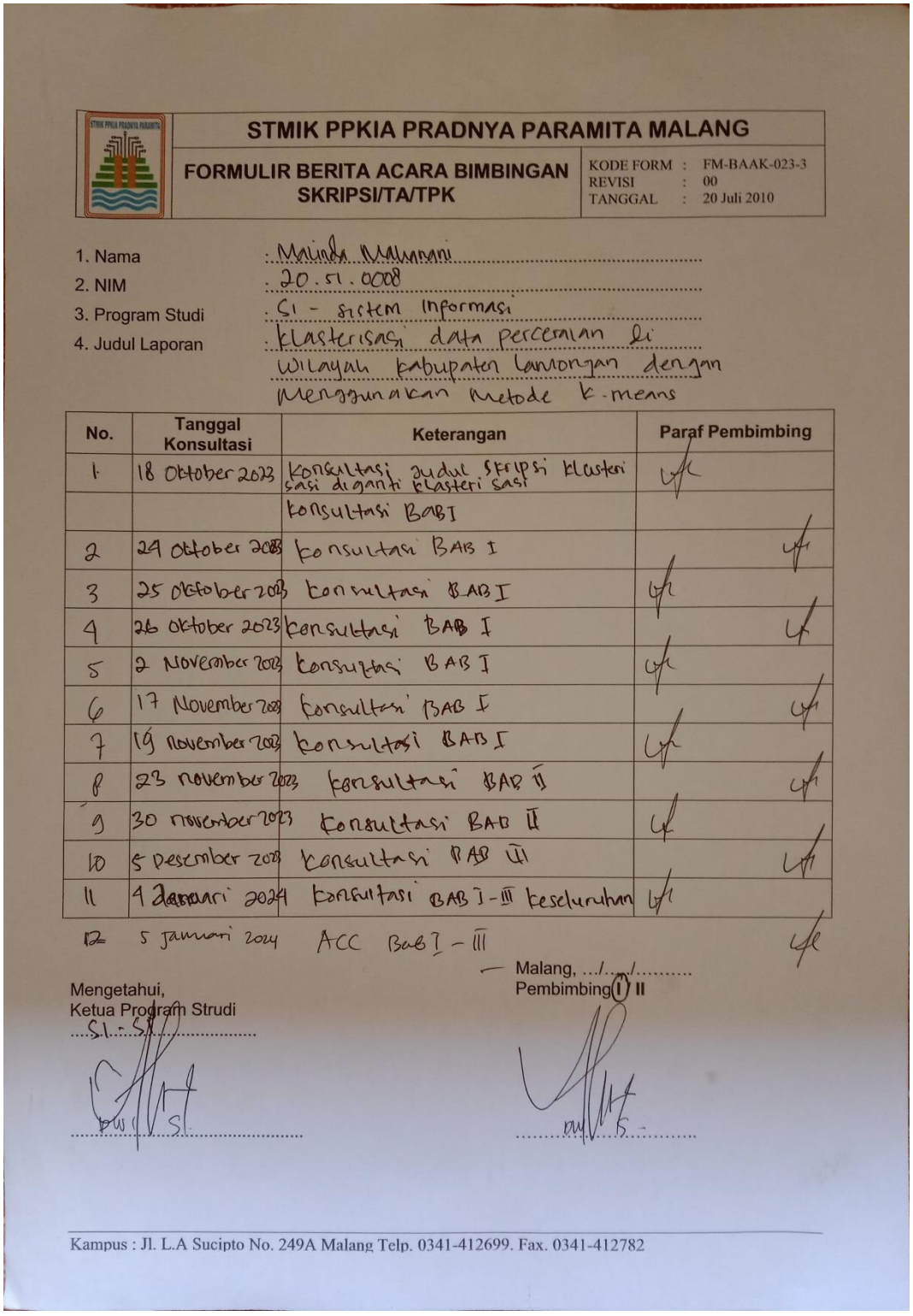
uci, d. r., Nuke , l. C., & Moch , F. (2022). **ANALISIS KASUS PERCERAIAN PADA PENGADILAN**. Jurnal IKRAITH-INFORMATIKA, 1.

Vada , A. E., & Arief , J. (2020). **KLUSTERISASI OPTIMAL DENGAN ELBOW METHOD UNTUK**. Dinamika Informatika,, 3.

RIWAYAT HIDUP

Malinda Mahanani dilahirkan di Lamongan, Jawa Timur pada tanggal 23 Maret 2002. Memulai pendidikan pada tahun 2007 di Taman Kanak-Kanak muslimat Badu Wanar. Melanjutkan pendidikan di MI Tarbiyatul Athfal Badu Wanar. Melanjutkan pendidikan di Sekolah Menengah Pertama di SMP Raden Patah Gempol Pading. Dan menyelesaikan Sekolah Menengah Atas pada tahun 2020 di SMA Hasyim Asyari 1 Pucuk. Setelah lulus SMA, melanjutkan pendidikan di perguruan tinggi STMIK PPKIA Pradnya Paramita Malang dan terdaftar sebagai mahasiswa jurusan S1-Sistem Informasi. Aktif mengikuti kegiatan ekstakurikuler di sekolah menengah pertama hingga sekolah menengah atas. Mengikuti kegiataan yang diadakan oleh kampus dan aktif di unit kegiatan mahasiswa sebagai sekretaris English Club pada tahun 2021 di kampus STIMATA Malang.

DAFTAR LAMPIRAN



STMIK PPKIA PRADNYA PARAMITA MALANG

FORMULIR BERITA ACARA BIMBINGAN SKRIPSI/TA/TPK

KODE FORM : FM-BAAK-023-3
REVISI : 00
TANGGAL : 20 Juli 2010

1. Nama : Malinda Maharani
2. NIM : 20.51.0008
3. Program Studi : S1 - sistem informasi
4. Judul Laporan : klasterisasi data perceraian di wilayah kabupaten lamongan dengan menggunakan metode k-means

No.	Tanggal Konsultasi	Keterangan	Paraf Pembimbing
1	18 Oktober 2023	konsultasi judul skripsi klusterisasi diganti klasterisasi	
		konsultasi BAB I	
2	24 Oktober 2023	konsultasi BAB I	
3	25 Oktober 2023	konsultasi BAB I	
4	26 Oktober 2023	konsultasi BAB I	
5	2 November 2023	konsultasi BAB I	
6	17 November 2023	konsultasi BAB I	
7	19 November 2023	konsultasi BAB I	
8	23 November 2023	konsultasi BAB II	
9	30 November 2023	konsultasi BAB II	
10	5 Desember 2023	konsultasi BAB III	
11	4 Januari 2024	konsultasi BAB I - III keseluruhan	
12	5 Januari 2024	ACC Bab I - III	

Malang, .../.../.....

Mengetahui,
Ketua Program Studi
S1 - SI

Pembimbing (1) II

Kampus : Jl. L.A Sucipto No. 249A Malang Telp. 0341-412699. Fax. 0341-412782