

Analisis Cluster Tingkat Perceraian Di Kota Malang Menggunakan Algoritma Agglomerative Hierarchical Clustering (Ahc)

Cholifatul Maulidia^{#1}, Rahayu Widayanti^{#2}, Dwi Safiroh Utsalina^{#3}

^{#1,2,3}Program Studi Sistem Informasi, STMIK Pradnya Paramita Malang

Korespondensi author *cholifatulmaulidiao@gmail.com^{#1}, rahayu@stimata.ac.id^{#2}, utsalina@stimata.ac.id^{#3}

Info Artikel

Diajukan: 10 Oktober 2025

Diterima: 10 Januari 2026

Diterbitkan: 26 Januari 2026

Keywords:

Divorce; Agglomerative Hierarchical Clustering; Average Linkage; Data Mining; CRISP-DM; Davies–Bouldin Index.

Kata Kunci:

Perceraian; Pengelompokan Hierarki Aglomeratif; Penggabungan Rata-Rata; Penambangan Data CRISP-DM; Davies–Bouldin Index.



Lisensi: cc-by-sa

Copyright © 2026 Cholifatul Maulidia, Rahayu Widayanti, Dwi Safiroh Utsalina

Abstract

The high divorce rate in Malang City is a social issue that requires data-driven analysis to support targeted policy formulation, leading this study to analyze the patterns and characteristics of divorce at the Malang City Religious Court using the Agglomerative Hierarchical Clustering (AHC) algorithm with the Average Linkage method. Utilizing a dataset of 5,891 cases from the 2022–2024 period—including attributes such as divorce type, ages of the parties, causal factors, and the district of marriage—the research followed the CRISP-DM framework and employed the Davies-Bouldin Index (DBI) to determine that the optimal number of clusters was 10 for 2022, 2023, and the combined data, and 11 for 2024. The findings reveal that the 20–30 age range is the group most vulnerable to divorce, primarily driven by constant disputes and arguments, with Kedungkandang, Sukun, and Blimbing identified as the districts with the highest divorce rates. Ultimately, this research serves as a foundation for relevant authorities to design more targeted divorce prevention programs based on specific regional and demographic characteristics.

Abstrak

Tingkat perceraian yang tinggi di Kota Malang merupakan isu sosial yang membutuhkan analisis berbasis data untuk mendukung perumusan kebijakan yang tepat sasaran. Studi ini bertujuan untuk menganalisis pola dan karakteristik perceraian di Pengadilan Agama Kota Malang menggunakan algoritma Agglomerative Hierarchical Clustering (AHC) dengan metode Average Linkage. Dengan menggunakan dataset sebanyak 5.891 kasus dari periode 2022–2024—termasuk atribut seperti jenis perceraian, usia pihak-pihak yang terlibat, faktor penyebab, dan kecamatan tempat pernikahan—penelitian ini mengikuti kerangka kerja CRISP-DM dan menggunakan Indeks Davies-Bouldin (DBI) untuk menentukan bahwa jumlah cluster optimal adalah 10 untuk data 2022, 2023, dan data gabungan, serta 11 untuk data 2024. Hasil penelitian menunjukkan bahwa kelompok usia 20–30 tahun merupakan kelompok yang paling rentan terhadap perceraian, terutama disebabkan oleh perselisihan dan pertengkaran yang terus-menerus, dengan Kedungkandang, Sukun, dan Blimbing diidentifikasi sebagai kecamatan dengan tingkat perceraian tertinggi. Pada akhirnya, penelitian ini berfungsi sebagai dasar bagi pihak berwenang terkait untuk merancang program pencegahan perceraian yang lebih tepat sasaran berdasarkan karakteristik regional dan demografis tertentu.

Cara mensitasi artikel:

C. Maulidia, R. Widayanti, D. S. Utsalina. "Analisis Cluster Tingkat Perceraian Di Kota Malang Menggunakan Algoritma Agglomerative Hierarchical Clustering (Ahc)" *DINAMIKA DOTCOM: Jurnal Pengembangan Manajemen Informatika & Komputer*, Vol. 17 No.1, pp 24-35, Januari 2026

PENDAHULUAN

Perceraian merupakan salah satu isu sosial yang mengalami peningkatan signifikan di berbagai wilayah Indonesia dalam beberapa tahun terakhir. Peningkatan angka perceraian tidak hanya berdampak pada stabilitas keluarga, tetapi juga pada aspek psikologis anak, ekonomi rumah tangga, serta harmonisasi sosial masyarakat [1], [2]. Kota Malang menjadi salah satu daerah dengan jumlah kasus perceraian yang tinggi, baik untuk jenis

cerai gugat maupun cerai talak. Faktor-faktor penyebab seperti Ekonomi, perselisihan dan pertengkaran terus menerus, ketidakharmonisan, menjadi pemicu utama meningkatnya angka perceraian hampir setiap tahun [3]. Kondisi ini menunjukkan perlunya analisis yang lebih komprehensif untuk memahami pola perceraian berbasis data secara akurat.

Pengadilan Agama Kota Malang telah menyediakan data statistik perceraian, namun pengelolaan data masih cenderung bersifat manual sehingga

menyulitkan proses identifikasi pola secara cepat dan terstruktur [4]. Oleh karena itu, penerapan metode analisis berbasis data mining sangat diperlukan untuk membantu pemerintah dan lembaga terkait dalam memahami tren perceraian secara mendalam. Data mining telah menjadi pendekatan yang efektif dalam menemukan pola tersembunyi dari dataset berukuran besar dan digunakan dalam berbagai penelitian sosial, kesehatan, maupun pendidikan. Salah satu teknik yang banyak digunakan adalah *clustering*, yaitu metode pengelompokan yang membagi objek ke dalam kelompok berdasarkan tingkat kemiripan tertentu.

Agglomerative Hierarchical Clustering

AHC adalah pendekatan bottom-up yang menggabungkan data berdasarkan kedekatan jarak antar objek. Pemilihan metode linkage mempengaruhi hasil cluster. Tiga metode yang sering digunakan adalah: *Single Linkage*, *Complete Linkage*, dan *Average Linkage* [6].

Dalam penelitian ini digunakan metode Average Linkage karena mampu menghasilkan cluster yang lebih stabil dan mengurangi risiko terjadinya chaining effect seperti pada Single Linkage. Metode ini menghitung jarak rata-rata antar seluruh pasangan data dalam dua cluster yang dibandingkan [7].

Davies–Bouldin Index (DBI)

DBI adalah metrik yang umum digunakan untuk mengevaluasi kualitas cluster. Indeks ini mengukur rasio antara jarak intra-cluster dan inter-cluster. Semakin kecil nilai DBI, semakin baik kualitas cluster yang terbentuk. DBI banyak digunakan dalam penelitian clustering karena mampu memberikan pengukuran objektif terhadap kualitas hasil pengelompokan [8].

Perceraian di Indonesia

Perceraian merupakan salah satu persoalan sosial yang terus meningkat dari tahun ke tahun. Berbagai penelitian menunjukkan bahwa perceraian banyak dipengaruhi oleh faktor internal maupun eksternal seperti perselisihan, kekerasan rumah tangga, masalah ekonomi, ketidakharmonisan, serta perubahan perilaku pasangan

[1]. Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa angka perceraian di Jawa Timur termasuk yang tertinggi di Indonesia, dan Kota Malang menjadi salah satu penyumbang utama dari peningkatan tersebut [13]. Kondisi ini menegaskan perlunya analisis berbasis data untuk melihat pola perceraian secara lebih objektif dan komprehensif.

METODE

Metode penelitian yang diterapkan sebagai solusi dalam penelitian ini mengikuti framework CRISP-DM (*Cross Industry Standard Process for Data Mining*), dengan siklus seperti dalam Gambar 1. CRISP-DM digunakan sebagai pendekatan penelitian karena menyediakan kerangka kerja yang sistematis dan terstruktur dalam menangani proses pengolahan data yang kompleks mulai dari tahap pemahaman bisnis hingga implementasi solusi berbasis data. Dalam konteks penelitian ini, CRISP-DM memungkinkan pengolahan data sosial secara iteratif, sehingga pemahaman terhadap data dan model dapat terus diperbarui seiring dengan meningkatnya pemahaman terhadap permasalahan sosial yang diteliti. CRISP-DM juga memberikan fleksibilitas dalam menyesuaikan strategi atau kebijakan pada tahap selanjutnya.



Gambar 1. *Framework CRISP-DM*

Data Understanding

Dataset yang digunakan dalam penelitian ini mencakup informasi data perceraian dari Pengadilan Agama Kota Malang 3 tahun terakhir yaitu pada tahun 2022-2024. Data ini meliputi atribut atribut berikut:

1. Jenis perceraian : Cerai Talak dan Cerai Gugat.

2. Umur Penggugat (p): Usia dari pihak yang mengajukan perceraian.
3. Umur Tergugat (t): Usia dari pihak yang digugat.
4. Penyebab perceraian: Alasan utama perceraian, seperti perselisihan, KDRT, masalah ekonomi.
5. Kecamatan tempat nikah (KUA): Kecamatan tempat menikah.

Dataset ini terdiri dari total 2.195 perkara pada tahun 2022, 1.990 perkara pada tahun 2023 dan 1.706 perkara pada tahun 2024, yang disimpan dalam format XLSX dan diolah menggunakan Microsoft Excel.

Data Preparation

Tahap ini bertujuan untuk mempersiapkan data agar siap digunakan dalam proses clustering. Langkah-langkah yang dilakukan dalam tahap ini meliputi:

- Pembersihan data: Menghapus fitur atau kolom yang tidak digunakan.
- Penyaringan data: Data yang berasal dari 'KUA Kecamatan' dengan nilai 'Luar Kota Malang dan Kota Batu' dihapus.
- konversi tipe data: Konversi ke dalam tipe data numerik, pada kolom 'umur P' dan 'umur T'.
- Transformasi Atribut Kategorikal: Pengkodean variable kategorikal yang merupakan variable kategorikal diubah menjadi bentuk numerik menggunakan Teknik One-Hot Encoding, agar atribut tersebut dapat diikutsertakan dalam perhitungan jarak antar data pada algoritma AHC.
- Normalisasi dan transformasi ini penting dilakukan karena algoritma AHC sangat bergantung pada nilai jarak antar data untuk menentukan struktur dendrogram. Dengan memastikan setiap atribut berada dalam skala yang setara dan representasi data yang seragam, hasil clustering yang dihasilkan menjadi lebih objektif dan representatif terhadap karakteristik data sebenarnya.

Evaluation

Evaluasi model clustering dilakukan untuk menilai kualitas dan validitas hasil Hierarchical Clustering (AHC).

Evaluasi ini menjadi tahap penting dalam proses analisis data, guna memastikan bahwa pola-pola yang terbentuk benar-benar mencerminkan struktur alami dari data yang dianalisis, dalam hal ini adalah data perceraian di Kota Malang. Dalam penelitian ini, pengukuran dilakukan menggunakan *Davies-Bouldin Index* (DBI). DBI digunakan untuk menilai seberapa baik suatu hasil *clustering*, dengan memperhatikan jarak rata-rata antar cluster dan kedekatan antar anggota dalam satu cluster. Agar DBI menghasilkan nilai terbaik maka sebisa mungkin Si harus bernilai sekecil mungkin dan M_{ij} harus bernilai sebesar mungkin. Semakin rendah nilai DBI yang dihasilkan, maka semakin baik kualitas *clustering*-nya. Nilai DBI juga digunakan untuk menentukan titik potong (*cut-off*) yang optimal dalam struktur *dendrogram*, sekaligus membantu dalam menentukan jumlah cluster terbaik.

Tahapan dalam melakukan perhitungan DBI Adalah sebagai berikut:

1. *Sum of Square Within Cluster* (SSW)

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Untuk mengetahui kohesi dalam sebuah *cluster* ke-*i* Adalah dengan menghitung *Sum of Square Within Cluster* (SSW).

2. *Sum of Square Between Cluster* (SSB)

$$SSB_{ij} = d(c_i, c_j)$$

Perhitungan *Sum of Square Between Cluster* (SSB) bertujuan untuk mengetahui jarak antar *cluster*.

3. *Ratio* (Rasio)

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*.

4. Davies Bouldin Index

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{ij})$$

Nilai rasio yang di peroleh dari persamaan rasio digunakan untuk mencari nilai DBI.

Selain evaluasi numerik menggunakan DBI, dilakukan pula analisis visual terhadap struktur dendrogram yang dihasilkan oleh AHC. Dendrogram dianalisis dengan mengamati perubahan jarak antar cluster, khususnya pada titik-titik di mana jarak mulai meningkat secara signifikan. Titik tersebut menjadi indikasi logis untuk memotong dendrogram, sehingga jumlah cluster yang diperoleh dapat dianggap paling representatif terhadap data yang dianalisis.

Dengan demikian, hasil pengelompokan data perceraian dapat dijadikan dasar yang valid untuk analisis lebih lanjut maupun perumusan kebijakan pencegahan perceraian yang lebih tepat sasaran.

HASIL DAN PEMBAHASAN

1) Data Sebelum Praproses

Penggabungan data dilakukan untuk periode tahun 2022 hingga 2024 dengan menambahkan field baru bernama tahun. Penambahan field ini merupakan bagian dari tahap praproses data guna mempermudah proses analisis selanjutnya.

```
print(f"data sebelum preprocessing", len(ds))
data sebelum preprocessing 5891
```

Gambar 2. Total data sebelum praproses

```
datasets = [
    (os.path.join(data_dir, 'data_perceraian_2022.xlsx'), '2022'),
    (os.path.join(data_dir, 'data_perceraian_2023.xlsx'), '2023'),
    (os.path.join(data_dir, 'data_perceraian_2024.xlsx'), '2024'),
]

dataframes = []
for file, tahun in datasets:
    if os.path.exists(file):
        df = pd.read_excel(file, engine='openpyxl')
        df['tahun'] = tahun
        dataframes.append(df)
    else:
        print(f'File not found: {file}')

dataframes = [
    df for df in dataframes
]

data_all = pd.concat(dataframes, ignore_index=True)

rename_columns = {}
if 'Umur Penggugat' in data_all.columns:
    rename_columns['Umur Penggugat'] = 'Umur P'
if 'Umur Tergugat' in data_all.columns:
    rename_columns['Umur Tergugat'] = 'Umur T'

data_all.rename(columns=rename_columns, inplace=True)

data_all.to_excel('data_perceraian_gabungan.xlsx', index=False, engine='openpyxl')
print("Combined data saved to 'data_perceraian_gabungan.xlsx'")
else:
    print("No dataframes were loaded. Please check the file paths and ensure files exist.")
```

Gambar 3. Kode Penggabungan data tiap periode

Kode program pada gambar 3 bertujuan untuk menggabungkan data perceraian dari beberapa file Excel tahunan (2022, 2023, dan 2024) yang tersimpan dalam folder /content/datasets Semua data yang berhasil dibaca kemudian digabungkan menjadi satu DataFrame menggunakan pd.concat() memeriksa keberadaan kolom 'Umur Penggugat' dan 'Umur Tergugat', lalu mengganti nama kolom tersebut menjadi 'Umur P' dan 'Umur T' agar lebih singkat dan seragam.

| No | Jenis Perkara | Umur P | Umur T | Faktor Penyebab | KUA Kecamatan | tahun | |
|------|---------------|-------------|--------|-----------------|---|--------------------------------|------|
| 0 | 1 | Cerai Talak | 53 | 56 | Perselisihan dan pertengkaran terus menerus | Luar Kota Malang dan Kota Batu | 2022 |
| 1 | 2 | Cerai Talak | 21 | 25 | Ekonomi | Sukun | 2022 |
| 2 | 3 | Cerai Talak | 21 | 25 | Perselisihan dan pertengkaran terus menerus | Lowokwaru | 2022 |
| 3 | 4 | Cerai Gugat | 21 | 25 | Perselisihan dan pertengkaran terus menerus | Luar Kota Malang dan Kota Batu | 2022 |
| 4 | 5 | Cerai Gugat | 44 | 49 | Perselisihan dan pertengkaran terus menerus | Lowokwaru | 2022 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 5886 | 1702 | Cerai Gugat | 40 | 28 | Perselisihan dan pertengkaran terus menerus | Sukun | 2024 |
| 5887 | 1703 | Cerai Talak | 39 | 25 | Perselisihan dan pertengkaran terus menerus | Kedungkandang | 2024 |
| 5888 | 1704 | Cerai Gugat | 35 | 31 | Ekonomi | Kedungkandang | 2024 |
| 5889 | 1705 | Cerai Gugat | 40 | 29 | Perselisihan dan pertengkaran terus menerus | Lowokwaru | 2024 |
| 5890 | 1706 | Cerai Gugat | 38 | 36 | Perselisihan dan pertengkaran terus menerus | Kedungkandang | 2024 |

Gambar 4. Hasil penggabungan data tiap periode

Gambar 4 menampilkan sebagian data hasil pengumpulan dari Pengadilan Agama Kota Malang yang mencatat perkara perceraian selama periode tahun 2022 hingga 2024. Data disajikan dalam bentuk tabel yang terdiri atas beberapa atribut, antara lain nomor, jenis perceraian, umur-p, umur-t, faktor penyebab perceraian, kecamatan tempat nikah (KUA), serta tahun terjadinya perkara.

2) Data Praproses

Pada tahap ini, dilakukan serangkaian proses pembersihan dan transformasi data untuk mempersiapkan data agar dapat digunakan dalam analisis lebih lanjut.

```
# praproses
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.compose import ColumnTransformer

ds = ds.drop(['No'], axis=1)

ds = ds[ds['KUA Kecamatan'] != 'Luar Kota Malang dan Kota Batu']
ds['Umur P'] = pd.to_numeric(ds['Umur P'], errors='coerce')
ds['Umur T'] = pd.to_numeric(ds['Umur T'], errors='coerce')

ds = ds.dropna(subset=['Umur P'])
ds = ds.dropna(subset=['Umur T'])

ds['Umur P'] = ds['Umur P'].astype(int)
ds['Umur T'] = ds['Umur T'].astype(int)

categorical_features = ['Faktor Penyebab', 'KUA Kecamatan']

preprocessor = ColumnTransformer([
    ('cat', OneHotEncoder(), categorical_features),
])

X_encoded = preprocessor.fit_transform(ds)
X_encoded = X_encoded.toarray() if hasattr(X_encoded, 'toarray') else X_encoded

print("data setelah preprocessing", len(ds))
data setelah preprocessing 4393
```

Gambar 5. Kode Praproses data

Gambar 5 menunjukkan kode yang digunakan dalam tahap praproses data sebelum dilakukan analisis lebih

lanjut. Tahapan ini mencakup proses pembersihan data, penghapusan nilai yang hilang atau tidak relevan, serta transformasi data ke dalam format yang sesuai untuk analisis. Selain itu, pada tahap ini juga dilakukan encoding terhadap data kategorikal agar dapat diolah oleh algoritma machine learning. Praproses data merupakan langkah penting untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan representatif, sehingga hasil analisis dapat memberikan informasi yang akurat dan dapat diinterpretasikan dengan baik. Proses ini menghasilkan total 4.393 data yang telah bersih dan siap digunakan. Langkah-langkah yang dilakukan antara lain:

- Penghapusan kolom yang tidak relevan:
Kolom 'No' dihapus karena tidak memiliki pengaruh terhadap proses analisis dan pemodelan.
- Penyaringan data:
Data yang berasal dari 'KUA Kecamatan' dengan nilai 'Luar Kota Malang dan Kota Batu' dihapus, fokus analisis hanya pada wilayah Kota Malang.
- Pembersihan dan konversi tipe data:
Kolom 'Umur P' (umur penggugat) dan 'Umur T' (umur tergugat) dikonversi ke dalam tipe numerik. Nilai-nilai yang tidak dapat dikonversi (misalnya berupa teks atau kosong) akan dikonversi menjadi NaN. Selanjutnya, baris yang memiliki nilai kosong pada kolom-kolom tersebut dihapus, dan tipe data dikonversi ke integer agar siap digunakan dalam analisis.
- Pengkodean variable kategorikal:
Kolom 'Faktor Penyebab' dan 'KUA Kecamatan', yang merupakan variabel kategorikal, diubah menjadi bentuk numerik menggunakan teknik *One-Hot Encoding*. Teknik ini mengubah setiap kategori menjadi kolom biner (0 atau 1), sehingga model dapat memproses informasi kategorikal dengan baik.

```

Hasil OneHot Encoding:
Faktor Penyebab_Cacat badan Faktor Penyebab_Dihukum penjara \
0 0,0 0,0
1 0,0 0,0
2 0,0 0,0
3 0,0 0,0
4 0,0 0,0
...
4388 0,0 0,0
4389 0,0 0,0
4390 0,0 0,0
4391 0,0 0,0
4392 0,0 0,0

Faktor Penyebab_Ekonomi Faktor Penyebab_Judi Faktor Penyebab_KDRT \
0 1,0 0,0 0,0
1 0,0 0,0 0,0
2 0,0 0,0 0,0
3 0,0 0,0 0,0
4 0,0 0,0 0,0
...
4388 0,0 0,0 0,0
4389 0,0 0,0 0,0
4390 1,0 0,0 0,0
4391 0,0 0,0 0,0
4392 0,0 0,0 0,0
    
```

Gambar 6. Hasil One Hot Encoding

Berdasarkan hasil One-Hot Encoding yang ditampilkan pada gambar 6, dapat dilihat bahwa hasil dari proses *One-Hot Encoding* yang diterapkan pada sebuah dataset. *One-Hot Encoding* adalah teknik yang digunakan untuk mengubah data kategorikal menjadi format numerik biner, di mana setiap kategori diwakili oleh sebuah vektor biner. Dalam vektor ini, hanya satu elemen yang bernilai 1 (menunjukkan keberadaan kategori tersebut), sementara elemen lainnya bernilai 0.

| Jenis Perkara | Umur P | Umur T | Faktor Penyebab | KUA Kecamatan tahun |
|---------------|-------------|--------|-----------------|--|
| 1 | Cerai Tatak | 21 | 25 | Ekonomi Sukun 2022 |
| 2 | Cerai Tatak | 21 | 25 | Perselisihan dan pertengkararan terus menerus Lowokasaru 2022 |
| 4 | Cerai Gugat | 44 | 49 | Perselisihan dan pertengkararan terus menerus Lowokasaru 2022 |
| 6 | Cerai Gugat | 21 | 25 | Perselisihan dan pertengkararan terus menerus Kligen 2022 |
| 7 | Cerai Gugat | 21 | 25 | Perselisihan dan pertengkararan terus menerus Junrejo 2022 |
| ... | ... | ... | ... | ... |
| 6886 | Cerai Gugat | 40 | 28 | Perselisihan dan pertengkararan terus menerus Sukun 2024 |
| 6887 | Cerai Tatak | 39 | 25 | Perselisihan dan pertengkararan terus menerus Kedungkandang 2024 |
| 6888 | Cerai Gugat | 35 | 31 | Ekonomi Kedungkandang 2024 |
| 6889 | Cerai Gugat | 40 | 29 | Perselisihan dan pertengkararan terus menerus Lowokasaru 2024 |
| 6890 | Cerai Gugat | 38 | 36 | Perselisihan dan pertengkararan terus menerus Kedungkandang 2024 |

4393 rows x 6 columns

Gambar 7 Hasil Data setelah dipraproses

Gambar 7 merupakan hasil dari proses *pra-pemrosesan data* yang dilakukan terhadap data perkara perceraian dari Pengadilan Agama Kota Malang dalam kurun waktu tahun 2022 hingga 2024. Tahapan *pra-pemrosesan* ini mencakup pembersihan data dari nilai-nilai yang tidak relevan atau tidak lengkap, serta standarisasi format data agar siap digunakan dalam proses analisis lebih lanjut. Setelah dilakukan *pra-pemrosesan*, data yang ditampilkan mencakup beberapa atribut penting, antara lain No (nomor data), Jenis Perceraian, Umur Penggugat (Umur P), Umur Tergugat (Umur T), Faktor Penyebab, KUA Kecamatan tempat pernikahan dilakukan, dan Tahun. Atribut Jenis Perceraian menunjukkan

kategori perceraian, yaitu Cerai Talak dan Cerai Gugat, yang masing-masing mengindikasikan pihak yang mengajukan perceraian (suami atau istri). Atribut Umur P dan Umur T telah dikonversi ke bentuk numerik dan telah dibersihkan dari nilai-nilai yang tidak logis. Sedangkan Faktor Penyebab telah disederhanakan dan dikategorikan sesuai dengan klasifikasi umum, seperti *Perselisihan dan pertengkaran terus menerus* dan *Ekonomi*. Proses ini dilakukan untuk memudahkan transformasi data ke dalam bentuk numerik yang dibutuhkan dalam analisis lanjutan, seperti teknik *one-hot encoding*.

3) Modeling

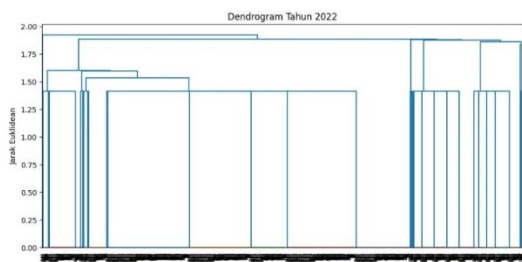
4) Pemisahan Data Pertahun

Dataset dibagi berdasarkan nilai pada kolom tahun. Setiap subset data mewakili data perceraian pada tahun tertentu.

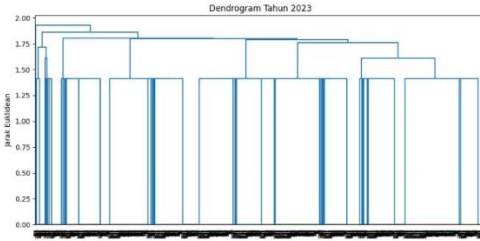
```
unique_years = sorted(ds['tahun'].unique())
for tahun in unique_years:
    subset = ds[ds['tahun'] == tahun]
    if len(subset) < 3:
        print(f"[Tahun {tahun}] Tidak cukup data untuk clustering ({len(subset)} data).")
        continue
```

Gambar 8 Kode pemisahan data tiap periode

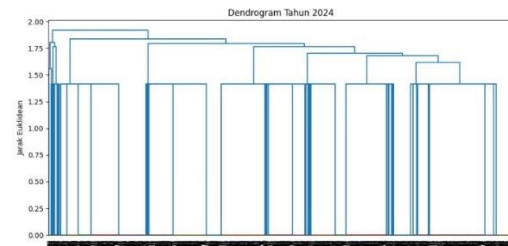
Kode pada Gambar 8 digunakan untuk memisahkan data berdasarkan tahun, yang diambil dari kolom 'tahun' dalam dataset ds. Tujuan dari proses ini adalah untuk melakukan visualisasi dendrogram data per tahun (periode) secara terpisah.



Gambar 9 Hasil Pemisahan data tahun 2022



Gambar 10 Hasil Pemisahan data tahun 2023



Gambar 11 Hasil Pemisahan data tahun 2024

Gambar 9,10,11 menampilkan hasil visualisasi dendrogram untuk data perceraian pertahun. Dendrogram merupakan representasi grafis dari proses hierarchical clustering yang menunjukkan bagaimana data dikelompokkan berdasarkan tingkat kemiripannya.

Penerapan *Agglomerative Hierarchical Clustering*

Untuk masing-masing subset tahunan yang memiliki jumlah data 3, dilakukan proses *Hierarchical Clustering* menggunakan metode *Average Linkage*. Metode ini dipilih karena memperhitungkan jarak rata-rata antar anggota *cluster*, sehingga menghasilkan struktur *cluster* yang lebih seimbang.

```
X_encoded = preprocessor.fit_transform(subset)
X_encoded = X_encoded.toarray() if hasattr(X_encoded, 'toarray') else X_encoded
Linkage_matrix = linkage(X_encoded, method='average')
```

Gambar 12 Kode penerapan Hierarchical Clustering

Gambar 12 kode penerapan *Hierarchical Clustering* data yang telah difilter sebelumnya, yang disimpan dalam variabel subset, diubah formatnya menggunakan sebuah objek preprocessor yang sudah disiapkan. Proses ini, yang disebut *fit_transform*, berfungsi untuk mengubah data mentah, seperti data teks atau data numerik yang belum terstandarisasi, menjadi

format numerik yang siap untuk analisis. Hasil dari proses ini kemudian disimpan dalam variabel `X_encoded`.

Langkah kedua adalah memastikan bahwa data `X_encoded` berada dalam format yang benar. Kode ini memeriksa apakah objek `X_encoded` memiliki metode untuk mengubahnya menjadi sebuah *array*. Jika ada, objek tersebut diubah menjadi *array* padat; jika tidak, data dibiarkan dalam format aslinya. Hal ini penting karena algoritma yang akan digunakan selanjutnya memerlukan *input* data dalam format yang spesifik.

Langkah terakhir adalah menerapkan algoritma *hierarchical clustering* pada data yang sudah disiapkan tersebut. Dengan menggunakan fungsi `linkage()`, kode ini menghitung jarak antar data dan menggabungkannya menjadi *cluster* menggunakan metode 'average', di mana jarak rata-rata antara dua *cluster* yang akan digabungkan menjadi acuan. Hasil dari proses ini adalah sebuah matriks yang merekam bagaimana *cluster-cluster* tersebut digabungkan, yang kemudian disimpan dalam variabel `linkage_matrix`.

5) Pengujian DBI

Pengujian menggunakan *Davies-Bouldin Index (DBI)* dilakukan untuk menentukan jumlah *cluster* yang paling optimal. DBI digunakan karena mampu mengukur kualitas pemisahan antar *cluster* dan kekompakan data dalam masing-masing *cluster* secara internal. Semakin kecil nilai DBI, maka semakin baik hasil pengelompokannya.

Dalam penelitian ini, pengujian dilakukan dengan mencoba beberapa titik potong (jumlah *cluster* yang berbeda), untuk membandingkan skor DBI yang dihasilkan. Nilai DBI terkecil dianggap sebagai indikasi bahwa jumlah *cluster* tersebut merupakan yang paling sesuai untuk data yang dianalisis. “Semakin kecil nilai DBI maka semakin baik hasil pengclusterannya”

```
print("Titik Potong\tDBI\tJumlah Cluster")
for t in [1, 1.2, 1.3, 1.4,1.45,1.5,1.6,1.7,1.75,1.8,1.9,2.0]:
    labels = fcluster(linkage_matrix, t=t, criterion='distance')
    n_clusters = len(set(labels))

    if n_clusters <= 1:
        print(f" {t:.2f}\t\tDBI tidak dihitung (hanya 1 cluster)")
        continue

    dbi = davies_bouldin_score(X_encoded, labels)
    print(f" {t:.2f}\t\t{dbi:.4f}\t\t{n_clusters}")
```

Gambar 13 Kode Evaluasi DBI

Kode pada gambar 13 berfungsi sebagai skrip evaluasi otomatis untuk menemukan titik potong optimal pada dendrogram AHC. Dengan mengiterasi serangkaian nilai titik potong (*t*), program secara berulang memotong dendrogram untuk membentuk *cluster* baru. Untuk setiap konfigurasi *cluster* yang dihasilkan (selama jumlahnya lebih dari satu), kode menghitung *Davies-Bouldin Score (DBI)*, sebuah metrik yang menilai kualitas *clustering* di mana nilai yang lebih rendah menandakan *cluster* yang lebih baik dan terpisah. Hasil dari setiap iterasi kemudian dicetak dalam format tabel, memungkinkan pengguna untuk dengan mudah membandingkan dan memilih nilai *t* yang menghasilkan DBI terendah, yang dianggap sebagai parameter terbaik untuk model *clustering*.

Berikut ini adalah Tabel yang menunjukkan hasil pengujian DBI berdasarkan berbagai jumlah *cluster* yang diuji:

Tabel 1 Hasil pengujian dbi periode 2022

| Jumlah Cluster | 10 | 10 | 8 | 7 | 7 | 7 | 2 |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| Hasil DBI | 1.2173 | 1.2173 | 1.2449 | 1.2718 | 1.2718 | 1.2718 | 1.2914 |
| Titik Potong | 1.45 | 1.50 | 1.60 | 1.70 | 1.75 | 1.80 | 1.90 |

Tabel 1 hasil pengujian DBI periode 2022 menghasilkan DBI sebesar 1.2173 dan jumlah *cluster* optimal 10.

Tabel 2 Hasil pengujian dbi periode 2023

| Jumlah Cluster | 10 | 10 | 8 | 7 | 7 | 7 | 2 |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| Hasil DBI | 1.0457 | 1.0457 | 1.2449 | 1.2718 | 1.2718 | 1.2718 | 1.2914 |
| Titik Potong | 1.45 | 1.50 | 1.60 | 1.70 | 1.75 | 1.80 | 1.90 |

Tabel 2 hasil pengujian DBI periode 2023 menghasilkan DBI sebesar 1.0457 dan jumlah *cluster* optimal 10.

Tabel 3 Hasil pengujian dbi periode 2024

| Jumlah Cluster | 11 | 11 | 10 | 8 | 7 | 4 | 2 |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| Hasil DBI | 1.0584 | 1.0584 | 1.3336 | 1.3897 | 1.4698 | 1.7382 | 2.3174 |
| Titik Potong | 1.45 | 1.50 | 1.60 | 1.70 | 1.75 | 1.80 | 1.90 |

Tabel 3 hasil pengujian DBI periode 2024 menghasilkan DBI sebesar 1.0584 dan jumlah cluster optimal 11.

Tabel 4 Hasil pengujian dbi periode gabungan

| Jumlah Cluster | 10 | 10 | 8 | 6 | 5 | 5 | 2 |
|----------------|--------|--------|--------|--------|--------|--------|--------|
| Hasil DBI | 1.1242 | 1.1242 | 1.2217 | 1.4141 | 1.4954 | 1.4954 | 1.2818 |
| Titik Potong | 1.49 | 1.50 | 1.60 | 1.70 | 1.75 | 1.80 | 1.90 |

Tabel 4 hasil pengujian DBI periode 2024 menghasilkan DBI sebesar 1.1242 dan jumlah cluster optimal 10.

Visualisasi Dendrogram

Setelah di temukan titik potong pada pengujian DBI selanjutnya Dataset dibagi berdasarkan nilai pada kolom setiap periode dan dilakukan visualisasi menggunakan Dendrogram. Setiap subset data mewakili data perceraian pada tahun tertentu lalu dilakukan visualisasi.

Gambar 14 merupakan potongan kode untuk menampilkan visualisasi dendrogram periode 2022,2023 dan 2024.

```
plt.figure(figsize=(10, 5))
dendrogram(linkage_matrix)
plt.title(f'Dendrogram Tahun {int(tahun)}')
plt.xlabel('Cluster ke-')
plt.ylabel('Jarak Euklidean')
plt.tight_layout()
plt.show()
```

Gambar 14 Visualisasi Dendrogram

Gambar 14 merupakan potongan kode untuk menampilkan visualisasi dendrogram periode gabungan. Kode diawali dengan membuat sebuah kanvas gambar berukuran 10 x 5 inci, yang berfungsi sebagai tempat untuk menggambar plot. Kemudian, sebuah dendrogram digambar menggunakan linkage_matrix, yaitu matriks yang berisi informasi tentang penggabungan cluster yang dihasilkan dari proses hierarchical clustering sebelumnya. Setelah dendrogram terbentuk, kode menambahkan

beberapa elemen untuk memperjelas visualisasi. Judul plot diatur menjadi "Dendrogram Tahun [Tahun]", di mana [Tahun] diganti dengan nilai tahun yang sedang dianalisis. Sumbu horizontal diberi label "Cluster ke-", sementara sumbu vertikal diberi label "Jarak Euklidean", yang mengindikasikan bahwa ketinggian garis pada dendrogram menunjukkan jarak antara cluster yang digabungkan.

Kode gambar 15 secara otomatis menyesuaikan tata letak plot agar label dan elemen lainnya tidak tumpang tindih, dan menampilkan dendrogram tersebut ke layar.

```
linkage_matrix_all = linkage(x_all_encoded, method='average')
from scipy.cluster.hierarchy import fcluster
from sklearn.metrics import davies_bouldin_score
for t in [1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.75, 1.8, 1.9, 2.0]:
    labels = fcluster(linkage_matrix_all, t, criterion='distance')
    if len(set(labels)) > 1:
        dbi = davies_bouldin_score(x_all_encoded, labels)
        print(f'Potong pada t={t:.2f} - DBI = {dbi:.4f}, Cluster = {len(set(labels))}')
    else:
        print(f'Potong pada t={t:.2f} - Cannot calculate DBI, Cluster = {len(set(labels))}')

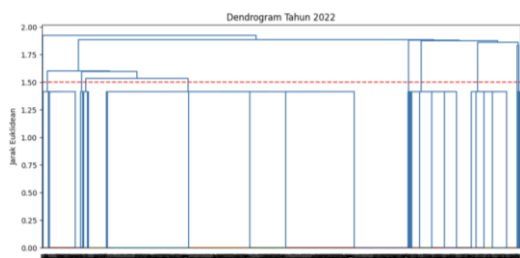
plt.figure(figsize=(15, 6))
dendrogram(linkage_matrix_all)
plt.title('Dendrogram Gabungan Semua Tahun')
plt.xlabel('Jarak Euklidean')
plt.axhline(y=1.50, color='r', linestyle='--')
plt.tight_layout()
plt.show()
print(len(x_all_encoded))
```

Gambar 15. Kode Visualisasi Dendrogram periode gabungan

Kode pada gambar 15 menampilkan proses analisis dan visualisasi hierarchical clustering pada data gabungan. Analisis diawali dengan membuat sebuah matriks keterkaitan (linkage_matrix_all) dari seluruh data yang telah diolah, menggunakan metode average. Matriks ini mendeskripsikan bagaimana setiap cluster digabungkan. Selanjutnya, untuk mengevaluasi kualitas clustering, kode ini menguji berbagai ambang batas jarak (threshold) dari 1.0 hingga 2.0. Untuk setiap ambang batas, cluster dibentuk, dan jika jumlah cluster lebih dari satu, kualitasnya dievaluasi menggunakan Davies-Bouldin Index (DBI). Nilai DBI, bersama dengan jumlah cluster yang terbentuk, dicetak untuk membantu menentukan ambang batas yang paling optimal.

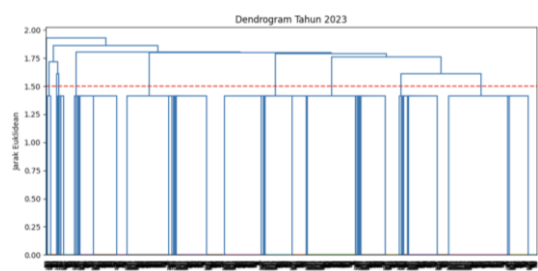
Untuk memvisualisasikan proses ini, sebuah dendrogram digambar. Dendrogram tersebut diberi judul "Dendrogram Gabungan Semua Tahun". Sumbu Y diberi label "Jarak Euklidean" untuk menunjukkan jarak antar cluster, dan sebuah garis horizontal berwarna merah pada

ketinggian 1.50 ditambahkan sebagai panduan untuk memotong *cluster*. Terakhir, kode ini mencetak jumlah total data yang digunakan dalam analisis.



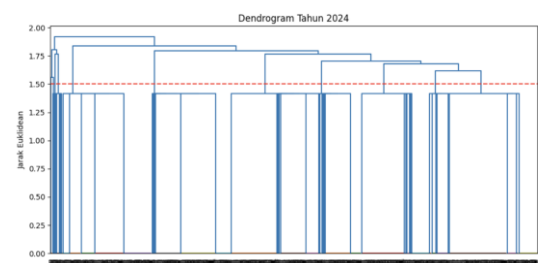
Gambar 16. Visualisasi Dendrogram periode 2022

Gambar 16 merupakan hasil visualisasi dendrogram periode 2022 dari hasil pengujian DBI periode 2022, pemotongan dendrogram pada ambang batas jarak sekitar 1.5, yang ditunjukkan oleh garis merah putus-putus, menghasilkan sebanyak 10 *cluster* optimal.



Gambar 17. Visualisasi Dendrogram periode 2023

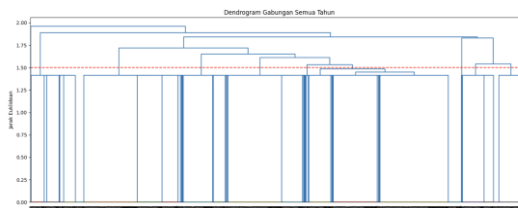
Gambar 17 merupakan hasil visualisasi dendrogram periode 2023 dari hasil pengujian DBI periode 2023, pemotongan dendrogram pada ambang batas jarak sekitar 1.5, yang ditunjukkan oleh garis merah putus-putus, menghasilkan sebanyak 10 *cluster* optimal.



Gambar 18. Visualisasi Dendrogram periode 2024

Gambar 18 merupakan hasil visualisasi dendrogram periode 2024 dari hasil pengujian DBI periode 2024, pemotongan dendrogram pada ambang batas jarak sekitar

1.5, yang ditunjukkan oleh garis merah putus-putus, menghasilkan sebanyak 11 *cluster* optimal.



Gambar 19. Visualisasi Dendrogram periode gabungan

Gambar 19 merupakan hasil visualisasi dendrogram periode gabungan dari hasil pengujian DBI periode gabungan, pemotongan dendrogram pada ambang batas jarak sekitar 1.5, yang ditunjukkan oleh garis merah putus-putus, menghasilkan sebanyak 10 *cluster* optimal.

Kesimpulan dari gambar Visualisasi dendrogram menunjukkan hasil *cluster* optimal yang diperoleh melalui metode *Agglomerative Hierarchical Clustering*. Sumbu vertikal menunjukkan nilai jarak antar data (dalam hal ini menggunakan metrik Euclidean), sedangkan sumbu horizontal merepresentasikan masing-masing objek data (kasus perceraian) yang dikelompokkan berdasarkan tingkat kemiripannya. Objek-objek yang bergabung pada tingkat yang lebih rendah menunjukkan kemiripan yang tinggi, sedangkan objek yang menyatu pada tingkat yang lebih tinggi memiliki perbedaan karakteristik yang lebih besar.

Evaluasi hasil DBI

Tabel 5. Hasil pengujian dbi periode gabungan

| Tahun | Titik Potong | Hasil DBI |
|----------|--------------|-----------|
| 2022 | 1.50 | 1.2173 |
| 2023 | 1.50 | 1.0457 |
| 2024 | 1.50 | 1.0584 |
| Gabungan | 1.49 | 1.1242 |

Tabel 5 menyajikan hasil perhitungan *Davies-Bouldin Index* (DBI) yang diperoleh untuk masing-masing periode tahun 2022, 2023, 2024, serta hasil gabungan dari ketiga periode tersebut. Nilai DBI digunakan untuk mengukur performa hasil *clustering*, di mana semakin kecil nilai DBI menunjukkan kualitas *clustering* yang semakin baik.

Pada tahun 2022, DBI yang dihasilkan adalah sebesar 1.2173 dengan titik potong 1.50. Pada tahun 2023, nilai DBI yang diperoleh mengalami penurunan menjadi 1.0457 pada titik potong 1.50. Tahun 2024 menunjukkan hasil DBI sebesar 1.0584 dengan titik potong 1.50. Sementara itu, pada data gabungan dari ketiga periode, nilai DBI tercatat sebesar 1.1242 dengan titik potong 1.49.

Deployment

Hasil dari proses *clustering* dimanfaatkan untuk memberikan rekomendasi dalam mengatasi faktor-faktor utama yang sering menjadi penyebab perceraian. Selain itu, hasil tersebut juga digunakan untuk mengidentifikasi wilayah dan rentang umur yang paling sering mengalami perceraian, sehingga dapat dijadikan sebagai dasar strategi penanganan oleh Pengadilan Agama Kota Malang.

Ringkasan Kluster Periode 2022:
total data 1589

| Cluster 2022 | Jumlah Data | Faktor Penyebab Dominan | Jumlah (Penyebab) | KUA Dominan | Jenis Perkara Dominan |
|--------------|-------------|---|-------------------|---------------|-----------------------|
| 0 | 1088 | Perselisihan dan pertengkaran terus menerus | 999 | Kedungkandang | Cerai Gugat |
| 1 | 12 | Dibuhuk penjara | 12 | Blimbing | Cerai Gugat |
| 2 | 8 | ABET | 8 | Sukun | Cerai Talak |
| 3 | 2 | Cacat badan | 2 | Lowokwaru | Cerai Gugat |
| 4 | 77 | Perselisihan dan pertengkaran terus menerus | 61 | Junrejo | Cerai Gugat |
| 5 | 152 | Meninggalkan salah satu pihak | 152 | Kedungkandang | Cerai Gugat |
| 6 | 2 | Murtad | 4 | Sukun | Cerai Gugat |
| 7 | 125 | Perselisihan dan pertengkaran terus menerus | 88 | Buntaji | Cerai Gugat |
| 8 | 18 | Perselisihan dan pertengkaran terus menerus | 7 | Batu | Cerai Gugat |
| 9 | 199 | Ekonomi | 199 | Sukun | Cerai Gugat |

Gambar 20. Distribusi dominan data di tiap cluster periode 2022

Berdasarkan gambar 20 "Ringkasan *cluster* Periode 2022", total data yang dianalisis adalah 1.589. Terdapat 10 *cluster*, yaitu *cluster* 0 hingga 9. *Cluster* dengan jumlah data terbanyak adalah *cluster* 0 dengan 1.000 data, di mana faktor penyebab dominan adalah "Perselisihan dan pertengkaran terus menerus" dan didominasi oleh KUA Kedungkandang. Jenis perceraian dominan pada *cluster* ini adalah Cerai Gugat. Selanjutnya, *cluster* 9 memiliki 199 data dengan faktor penyebab dominan "Ekonomi" dan didominasi oleh KUA Sukun, dengan jenis perceraian dominan "Cerai Gugat". *Cluster* 5 memiliki 152 data, dengan faktor penyebab dominan "Meninggalkan salah satu pihak" dan didominasi oleh KUA Kedungkandang, dengan jenis perceraian dominan "Cerai Gugat". *Cluster* dengan jumlah data terendah adalah *cluster* 3 dengan 2 data, dan faktor penyebab dominan "Cacat badan". *Cluster* 1, 2, 6, dan 8 juga memiliki jumlah data yang

rendah, dengan jenis perceraian dominan yang sama, yaitu "Cerai Gugat".

Ringkasan Kluster Periode 2023:
total data 1510

| Cluster 2023 | Jumlah Data | Faktor Penyebab Dominan | Jumlah (Penyebab) | KUA Dominan | Jenis Perkara Dominan |
|--------------|-------------|---|-------------------|---------------|-----------------------|
| 0 | 273 | Perselisihan dan pertengkaran terus menerus | 273 | Kedungkandang | Cerai Gugat |
| 1 | 299 | Perselisihan dan pertengkaran terus menerus | 129 | Sukun | Cerai Gugat |
| 2 | 283 | Perselisihan dan pertengkaran terus menerus | 149 | Blimbing | Cerai Gugat |
| 3 | 28 | Ekonomi | 19 | Junrejo | Cerai Gugat |
| 4 | 3 | Mabuk | 3 | Buntaji | Cerai Gugat |
| 5 | 255 | Perselisihan dan pertengkaran terus menerus | 134 | Lowokwaru | Cerai Gugat |
| 6 | 52 | Ekonomi | 24 | Buntaji | Cerai Gugat |
| 7 | 161 | Perselisihan dan pertengkaran terus menerus | 72 | Klojen | Cerai Gugat |
| 8 | 152 | Perselisihan dan pertengkaran terus menerus | 88 | Kedungkandang | Cerai Gugat |
| 9 | 4 | Meninggalkan salah satu pihak | 3 | Batu | Cerai Gugat |

Gambar 21. Distribusi dominan data di tiap cluster periode 2023

Berdasarkan gambar 21 "Ringkasan *Cluster* Periode 2023", total data yang dianalisis adalah 1.510. Terdapat 10 *cluster*, yaitu *cluster* 0 hingga 9. *Cluster* dengan jumlah data terbanyak adalah *cluster* 1 (299 data), *cluster* 2 (283 data), dan *cluster* 0 (273 data). Pada ketiga *cluster* tersebut, faktor penyebab dominan adalah "Perselisihan dan pertengkaran terus menerus", dan jenis perceraian dominannya adalah "Cerai Gugat". *Cluster* 5 juga memiliki jumlah data yang signifikan, yaitu 255 data, dengan faktor penyebab dominan "Perselisihan dan pertengkaran terus menerus" dan jenis perceraian "Cerai Gugat". *Cluster* dengan jumlah data terendah adalah *cluster* 4 (3 data) dengan faktor penyebab dominan "Mabuk", dan *cluster* 9 (4 data) dengan faktor penyebab dominan "Meninggalkan salah satu pihak". Meskipun faktor penyebabnya berbeda-beda, jenis perceraian dominan pada semua *cluster* adalah "Cerai Gugat".

Ringkasan Kluster Periode 2024:
total data 1294

| Cluster 2024 | Jumlah Data | Faktor Penyebab Dominan | Jumlah (Penyebab) | KUA Dominan | Jenis Perkara Dominan |
|--------------|-------------|---|-------------------|---------------|-----------------------|
| 0 | 236 | Perselisihan dan pertengkaran terus menerus | 236 | Kedungkandang | Cerai Gugat |
| 1 | 17 | Meninggalkan salah satu pihak | 7 | Buntaji | Cerai Gugat |
| 2 | 236 | Perselisihan dan pertengkaran terus menerus | 122 | Blimbing | Cerai Gugat |
| 3 | 4 | Meninggalkan salah satu pihak | 3 | Batu | Cerai Gugat |
| 4 | 52 | Meninggalkan salah satu pihak | 38 | Kedungkandang | Cerai Gugat |
| 5 | 236 | Meninggalkan salah satu pihak | 236 | Sukun | Cerai Gugat |
| 6 | 5 | Ekonomi | 5 | Buntaji | Cerai Gugat |
| 7 | 118 | Perselisihan dan pertengkaran terus menerus | 68 | Klojen | Cerai Gugat |
| 8 | 181 | Perselisihan dan pertengkaran terus menerus | 114 | Sukun | Cerai Gugat |
| 9 | 218 | Perselisihan dan pertengkaran terus menerus | 93 | Lowokwaru | Cerai Gugat |
| 10 | 7 | Meninggalkan salah satu pihak | 7 | Junrejo | Cerai Gugat |

Gambar 22. Distribusi dominan data di tiap cluster periode 2024

Berdasarkan gambar 22 "Ringkasan *cluster* Periode 2024", total data yang dianalisis adalah 1.294. Terdapat 11 *cluster*, yaitu *cluster* 0 hingga 10. *Cluster* 0, 2, 5, dan 9 memiliki jumlah data tertinggi, masing-masing 236, 236, 236, dan 210 data. Faktor penyebab dominan pada *cluster-cluster* ini bervariasi, seperti "Perselisihan dan pertengkaran terus menerus" pada *cluster* 0, 2, dan 9, serta "Ekonomi" pada *cluster* 5. Meskipun demikian, jenis perceraian dominan pada semua *cluster* adalah "Cerai Gugat".

Gugat". *Cluster* dengan jumlah data terendah adalah *cluster* 3 dengan 4 data, *cluster* 6 dengan 5 data, dan *cluster* 10 dengan 7 data, di mana faktor penyebab dominannya juga beragam.

Berikut ini gambar 23 Distribusi dominan data di tiap *cluster* periode 2022 – 2024 yang digunakan untuk uji validasi kesesuaian *cluster*:

Ringkasan Kluster Periode Gabungan:

| Cluster gabungan | Jumlah Data | Faktor Penyebab Dominan | Jumlah (Penyebab) | KUA Dominan | Jumlah (KUA) | Jenis Perkara Dominan |
|------------------|-------------|---|-------------------|---------------|--------------|-----------------------|
| 0 | 1347 | Perselisihan dan pertengkaran terus menerus | 1388 | Kedungkandang | 605 | Cerai Gugat |
| 1 | 690 | Perselisihan dan pertengkaran terus menerus | 474 | Blimbing | 508 | Cerai Gugat |
| 2 | 875 | Perselisihan dan pertengkaran terus menerus | 471 | Sukun | 875 | Cerai Gugat |
| 3 | 261 | Ekonomi | 288 | Kedungkandang | 261 | Cerai Gugat |
| 4 | 23 | Batu | 23 | Batu | 23 | Cerai Gugat |
| 5 | 465 | Meninggalkan salah satu pihak | 465 | Lowokwaru | 115 | Cerai Gugat |
| 6 | 395 | Perselisihan dan pertengkaran terus menerus | 759 | Klojen | 395 | Cerai Gugat |
| 7 | 6 | Mabuk | 6 | Bumiaji | 4 | Cerai Gugat |
| 8 | 323 | Ekonomi | 323 | Lowokwaru | 192 | Cerai Gugat |
| 9 | 8 | Murtad | 8 | Kedungkandang | 3 | Cerai Gugat |

Gambar 23. Distribusi dominan data di tiap cluster periode 2022-2024

Berdasarkan gambar 23 "Ringkasan *cluster* Periode Gabungan", total data yang dianalisis adalah 4.393. Terdapat 10 *cluster* yang terbentuk, yaitu *cluster* 0 hingga 9. *Cluster* dengan jumlah data terbanyak adalah *cluster* 0 dengan 1.347 data, di mana faktor penyebab dominan adalah "Perselisihan dan pertengkaran terus menerus" dan didominasi oleh KUA Kedungkandang. Jenis perceraian dominan pada *cluster* ini adalah "Cerai Gugat". *Cluster* 2 memiliki 875 data, dengan faktor penyebab dominan "Perselisihan dan pertengkaran terus menerus", dan didominasi oleh KUA Sukun. *Cluster* 1 memiliki 690 data, dengan faktor penyebab dominan "Perselisihan dan pertengkaran terus menerus" dan didominasi oleh KUA Blimbing. *Cluster* 5 memiliki 465 data dengan faktor penyebab dominan "Meninggalkan salah satu pihak" dan didominasi oleh KUA Lowokwaru. *Cluster* 6 memiliki 395 data, dengan faktor penyebab dominan "Perselisihan dan pertengkaran terus menerus" dan didominasi oleh KUA Klojen. *Cluster* 8 memiliki 323 data dengan faktor penyebab dominan "Ekonomi" dan didominasi oleh KUA Lowokwaru. *Cluster* dengan jumlah data terendah adalah *cluster* 7 (6 data) dengan faktor penyebab dominan "Mabuk", serta *cluster* 9 (8 data) dengan faktor penyebab dominan "Murtad". Meskipun faktor penyebabnya beragam, jenis perceraian dominan pada semua *cluster* adalah "Cerai Gugat".

Tabel 6 distribusi data *cluster* gabungan, faktor utama penyebab perceraian di berbagai pernikahan.

Tabel 6. Distribusi data cluster gabungan

| Cluster | KUA Kecamatan | Jumlah Data | Faktor Dominan | Persentase Faktor | Persentase KUA |
|---------|---------------|-------------|---|-------------------|----------------|
| 0 | Kedungkandang | 605 | Perselisihan dan pertengkaran terus menerus | 96.51% | 44,9% |
| 1 | Blimbing | 690 | Perselisihan dan pertengkaran terus menerus | 68.70 % | 100% |
| 2 | Sukun | 875 | Perselisihan dan pertengkaran terus menerus | 53.83% | 100% |
| 3 | Kedungkandang | 261 | Ekonomi | 78.16% | 100% |
| 4 | Batu | 23 | Perselisihan dan pertengkaran terus menerus | 82.61% | 100% |
| 5 | Lowokwaru | 115 | Meninggalkan salah satu pihak | 100.00 % | 24,7% |
| 6 | Klojen | 395 | Perselisihan dan pertengkaran terus menerus | 65.57 % | 100% |
| 7 | Bumiaji | 4 | Mabuk | 100% | 66,7% |
| 8 | Lowokwaru | 192 | Ekonomi | 100% | 59,4% |
| 9 | Kedungkandang | 3 | Murtad | 100% | 37,5% |

Berdasarkan tabel 6 Distribusi *cluster* data gabungan, dapat diidentifikasi beberapa pola dan karakteristik pada masing-masing *cluster*. *Cluster* 0, 1, 2, 4, dan 6 memiliki faktor dominan "Perselisihan dan pertengkaran terus menerus". *Cluster* 1, 2, dan 6 menunjukkan dominasi 100% pada KUA masing-masing (Blimbing, Sukun, dan Klojen). Sementara itu, *cluster* 0 dan 4 memiliki persentase KUA dominan yang lebih rendah. *cluster* 0 memiliki jumlah data terbanyak (605) dan persentase faktor dominan 96,51%.

Faktor dominan "Ekonomi" ditemukan pada *cluster* 3 dan 8. *Cluster* 3 didominasi oleh KUA Kedungkandang dan memiliki persentase KUA 100%. *Cluster* 8 didominasi oleh KUA Lowokwaru dengan persentase KUA 59,4%. Faktor dominan "Meninggalkan salah satu pihak" terdapat pada *cluster* 5, yang didominasi oleh KUA Lowokwaru. *Cluster* ini memiliki persentase faktor dominan 100%. Faktor dominan lainnya adalah "Mabuk" pada *cluster* 7, "Murtad" pada *cluster* 9, yang didominasi oleh KUA Bumiaji dan KUA Kedungkandang. *Cluster* 7, 8, dan 9 memiliki jumlah data yang sangat sedikit, yaitu 4, 192, dan 3 data.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini dilakukan dengan mengintegrasikan variabel jenis perceraian, usia pasangan, faktor penyebab, serta wilayah asal pernikahan (KUA). Hasil pemetaan menunjukkan konsistensi jumlah kelompok setiap tahunnya, data Gabungan (2022–2024) Menghasilkan 10 *cluster* dari total 4.393 kasus. Stabilitas jumlah *cluster* ini mengindikasikan belum adanya intervensi signifikan dari instansi terkait dalam menekan angka perceraian. Faktor utama penyebab perceraian tetap didominasi oleh perselisihan dan pertengkaran yang berlarut-larut, yang berakar pada masalah komunikasi dalam rumah tangga.

Secara keseluruhan, penggunaan *Hierarchical Clustering* berhasil mengungkap karakteristik spesifik di balik angka perceraian di Kota Malang. Temuan ini dapat menjadi landasan bagi pembuat kebijakan untuk menyusun program intervensi yang lebih presisi, tidak hanya sekadar edukasi umum, namun menasar akar permasalahan pada setiap kelompok masyarakat.

Saran

Penelitian selanjutnya disarankan untuk tidak hanya terbatas pada penggunaan algoritma *Agglomerative Hierarchical Clustering* dengan metode *Average Linkage*, tetapi juga mencoba membandingkan metode *clustering* lainnya seperti *K-Means Clustering*, *DBSCAN*, atau *Gaussian Mixture Model (GMM)*.

UCAPAN TERIMA KASIH

Peneliti menyampaikan terimakasih kepada para dosen pembimbing dan orang tua yang telah mendukung serta penyedia platform publikasi atas kemudahan aksesibilitas bagi para pembaca. Keberhasilan penelitian ini merupakan dari kolaborasi dan motivasi yang diberikan oleh berbagai pihak terkait. Terima kasih atas kepercayaan dan kesempatan yang telah diamanahkan kepada peneliti.

REFERENSI

- [1] Y. Sopyan, A. D. Lesmana and C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Building of Informatics, Technology and Science (BITS)*, 2022.
- [2] N. Veronika, P. C. Azhar and A. R. Sugma, "Dampak Perceraian Terhadap Psikologi Anak," *JBS (Jurnal Berbasis Sosial)*, pp. 30 - 37, 2022.
- [3] K. Kurniawan, "Jumlah Angka Perceraian di Kota Malang Masih Tinggi, Judi Jadi Salah Satu Penyebabnya," 2024. [Online]. Available: https://jatim.tribunnews.com/2024/11/29/jumlah-angka-perceraian-di-kota-malang-masih-tinggi-judi-jadi-salah-satu-penyebabnya#google_vignette.
- [4] "Laporan Perkara Perceraian," Malang, 2025.
- [5] B. M. Kumarahadi, H. Pratiwi and S. Subanti, "Penerapan Metode Hierarchical Clustering Untuk Pengelompokan Kota/Kabupaten Di Indonesia Berdasarkan Indikator Kemiskinan," *Jurnal TIKomSiN*, pp. 8-12, 2023.
- [6] S. and H. Wintolo, "SYSTEM FOR DETERMINING PUBLIC HEALTH LEVEL USING THE AGGLOMERATIVE HIERARCHICAL CLUSTERING METHOD," 2019.
- [7] G. Abdurrahman, "Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage," *JUSTINDO (Jurnal Sistem & Teknologi Informasi Indonesia)*, pp. 13-20, 2019.
- [8] A. K. B. Ginting, A. P. Harianja and S. P. Sipayung, "Efektivitas Metode Gap Statistic dan X-Means dalam Menentukan Jumlah Cluster Optimal pada K-Means Clustering," *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, pp. 133-139, 2024.
- [9] M. Mahanani, D. S. Utsalina and T. M. Akhriza, "REKOMENDASI SOLUSI ATAS KASUS PERCERAIAN DI KABUPATEN LAMONGAN BERBASIS K-MEANS CLUSTERING," *Seminar Nasional Teknologi Informasi dan Komunikasi (SeNTIK STI&K)*, pp. 489 - 496, 2024.
- [10] A. and T. Rima, "Data Mining penentuan jurusan Siswa Menggunakan Metode Agglomerative Hierarchical Clustering," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, pp. 876-880, April 2023.
- [11] Samsi and A. Mustafa, "Penerapan Algoritma AHC dalam mengelompokkan perkara perceraian," pp. 1-149, 2021.
- [12] H. Februariyanti and D. B. Santoso, "HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK PENGELOMPOKAN SKRIPSI MAHASISWA," *Prosiding SINTAK 2017*, pp. 33-40, 2017.
- [13] F. N. Hidayah, "GoodStats," 2023. [Online]. Available: <https://data.goodstats.id/statistic/5-faktor-tertinggi-penyebab-perceraian-di-indonesia-HLBgQ>.