

Implementasi *Data Mining Clustering* Dalam Mengelompokkan Kasus Perceraian di Provinsi Jawa Timur Menggunakan Algoritma K-Means

Saylendra Arga Wardani^{1*}, Nizar Bahri Al Varuq², Heru Teguh Santoso³

¹Program Studi Sistem Informasi, Universitas Nusantara PGRI Kediri, Indonesia

argasaylendra@gmail.com¹, nizarqwh8@gmail.com², herusants22@gmail.com³

Correspondence: duiswaute@udgi.ac.id¹

ABSTRAK

Fenomena perceraian di Provinsi Jawa Timur menunjukkan tren peningkatan setiap tahunnya, didorong oleh berbagai faktor kompleks seperti kondisi ekonomi, pendidikan, dan sosial budaya. Penelitian ini menggunakan pendekatan data mining dengan algoritma K-Means untuk mengelompokkan kasus perceraian berdasarkan kesamaan karakteristik. Data perceraian dari Badan Pusat Statistik tahun 2020-2022 dianalisis menggunakan metode clustering, dengan evaluasi kualitas hasil clustering dilakukan menggunakan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa model clustering dengan 7 klaster memberikan performa terbaik dengan nilai DBI terendah, yaitu 0,460. Setiap klaster merepresentasikan pola distribusi perceraian yang unik, yang dapat digunakan untuk mengidentifikasi kelompok masyarakat yang rentan terhadap perceraian. Temuan ini diharapkan dapat memberikan wawasan yang mendalam bagi pengambil kebijakan dalam merancang strategi pencegahan perceraian yang lebih tepat sasaran. Penelitian lebih lanjut disarankan untuk memperkaya analisis dengan variabel tambahan seperti kondisi ekonomi dan tingkat pendidikan, serta menggunakan data yang lebih terkini.

Kata Kunci

Perceraian; Clustering; Algoritma K-Means; Data mining; Davies-Bouldin Index;

ABSTRACT

The phenomenon of divorce in East Java Province shows an increasing trend every year, driven by various complex factors such as economic, educational, and socio-cultural conditions. This research uses a data mining approach with the K-Means algorithm to cluster divorce cases based on similar characteristics. Divorce data from the Central Bureau of Statistics for 2020-2022 was analyzed using the clustering method, with an evaluation of the quality of the clustering results carried out using the Davies-Bouldin Index (DBI). The results show that the clustering model with 7 clusters provides the best performance with the lowest DBI value, which is 0.460. Each cluster represents a unique divorce distribution pattern, which can be used to identify groups of people who are vulnerable to divorce. The findings are expected to provide deep insights for policy makers in designing more targeted divorce prevention strategies. Further research is recommended to enrich the analysis with additional variables such as economic conditions and education levels, as well as using more recent data.

Key Words

Divorce; Clustering; K-Means algorithm; Data mining; Davies-Bouldin Index;

Received: 26th December 2024

Accepted: 12th April 2025

Published: 30th June 2025

Citation: -

10.46510/jami.v6i1.324

ISSN 2722-4414 (p)/ 2722-4406 (e)

68

<https://journal.akb.ac.id/>

I. PENDAHULUAN

Salah satu masalah sosial yang semakin menjadi perhatian di Indonesia[1], termasuk di Provinsi Jawa Timur, adalah perceraian. Angka perceraian di Jawa Timur cenderung meningkat setiap tahunnya, menunjukkan kompleksitas masalah yang dihadapi masyarakat. Ini ditunjukkan oleh data dari Badan Pusat Statistik Provinsi Jawa Timur. Faktor-faktor seperti masalah ekonomi, ketidakcocokan dalam rumah tangga, pendidikan yang rendah, dan pengaruh sosial dan budaya juga berkontribusi pada peningkatan jumlah perceraian [2]. Fenomena ini memerlukan perhatian khusus, baik dari pemerintah maupun lembaga sosial, untuk mencari solusi yang dapat menekan angka perceraian[3].

Memahami faktor-faktor yang memengaruhi perceraian tidaklah mudah. Data perceraian yang tersedia sering kali sangat kompleks, dengan banyak variabel yang saling berkaitan[4]. Maka dari itu, diperlukan pendekatan analisis yang tepat guna menggali informasi dari data tersebut. Salah satu pendekatan yang dapat memberikan wawasan lebih mendalam adalah data mining, yang memungkinkan peneliti untuk mengeksplorasi dan mengekstrak pola-pola tersembunyi dalam kumpulan data yang besar[5]. Dengan ini, diharapkan analisis yang lebih terperinci terhadap faktor-faktor penyebab perceraian dapat dilakukan, sehingga langkah-langkah pencegahan yang tepat dapat diambil.

Di dalam data mining, metode clustering menjadi salah satu teknik yang efektif dalam mengelompokkan data ke dalam kelompok-kelompok berdasarkan kesamaan tertentu[6]. Teknik ini bertujuan untuk menemukan pola umum[7] di antara kasus-kasus perceraian, sehingga bisa dikelompokkan ke dalam beberapa kategori. Salah satu algoritma clustering yang populer dan efisien adalah algoritma K-Means[8]. Algoritma ini berfungsi untuk membagi data menjadi beberapa kluster yang memiliki karakteristik serupa[9]. Dalam konteks perceraian, K-Means dapat membantu mengidentifikasi pola perceraian berdasarkan berbagai variabel, seperti usia pasangan, latar belakang ekonomi, tingkat pendidikan, dan faktor lainnya.

Penelitian ini mengusulkan penerapan algoritma K-Means untuk mengelompokkan kasus-kasus perceraian yang terjadi di Provinsi Jawa Timur. Dengan menggunakan data perceraian yang tersedia, algoritma ini akan mengelompokkan kasus perceraian ke dalam beberapa kluster yang mewakili karakteristik yang serupa antar kasus. Hasil dari pengelompokan ini diharapkan dapat memberikan gambaran yang lebih jelas tentang tipe-tipe perceraian yang paling sering terjadi di Jawa Timur, serta faktor-faktor utama yang berkontribusi pada setiap kluster perceraian.

Selain itu, implementasi algoritma K-Means dalam konteks sosial seperti perceraian memiliki manfaat yang signifikan. Dengan adanya pengelompokan kasus perceraian, para pembuat kebijakan dapat lebih memahami profil kelompok-kelompok masyarakat yang rentan terhadap perceraian. Informasi ini akan menjadi landasan penting dalam perumusan kebijakan sosial, terutama dalam upaya pencegahan perceraian, konseling keluarga, serta program peningkatan kesejahteraan sosial.

Tujuan utama dari penelitian ini adalah untuk memberikan pemahaman yang lebih baik mengenai pola perceraian yang terjadi di Provinsi Jawa Timur dengan menggunakan pendekatan data mining. Melalui penerapan algoritma K-Means, diharapkan penelitian ini dapat membantu memetakan masalah perceraian secara lebih terstruktur dan memberikan solusi berbasis data untuk mengatasi masalah tersebut. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan ilmu pengetahuan, tetapi juga memiliki implikasi praktis yang dapat diimplementasikan oleh berbagai pihak dalam menanggulangi angka perceraian yang semakin tinggi.

II. METODE PENELITIAN

Penelitian ini menggunakan metode clustering; clustering adalah salah satu metode fungsi data mining yang menggunakan algoritma untuk mengelompokkan sejumlah data ke dalam kelompok-kelompok tertentu (cluster)[10]. Salah satu metode clustering non-hirarki, K-Means, berusaha mempartisi data saat ini ke dalam satu atau lebih cluster atau kelompok. Mampu mengelompokkan objek besar dan mempercepat proses pengelompokan adalah keuntungan menggunakan K-Means [11].

Data Mining

Data mining adalah suatu proses ekstraksi atau penggalian data dan informasi yang besar, yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting[16]. Data mining menggambarkan sebuah pengumpulan teknik-teknik dengan tujuan untuk menemukan pola-

pola yang tidak diketahui pada data yang telah dikumpulkan. Data mining merupakan bagian proses KDD (Knowledge Discovery in Databases), KDD secara umum juga dikenal sebagai pangkalan data.[17]

Tujuan utama dari data mining adalah mengidentifikasi pola, hubungan, dan tren yang tersembunyi dalam data untuk mendapatkan wawasan yang berharga mendukung pengambilan keputusan yang lebih baik.

Clustering

Cluster adalah sekumpulan data yang dikelompokkan berdasarkan kesamaan tertentu antara sebagian data, dengan perbedaan signifikan dari kelompok lainnya[12]. Berbeda dengan klasifikasi, clustering tidak berfokus pada prediksi nilai variabel objek, melainkan pada pengelompokan data. Algoritma clustering bertujuan untuk membagi informasi yang ada ke dalam kelompok-kelompok yang serupa. Semakin besar kemiripan data dalam satu cluster, semakin baik hasilnya, sementara perbedaan antar cluster akan menghasilkan nilai yang semakin kecil[13].

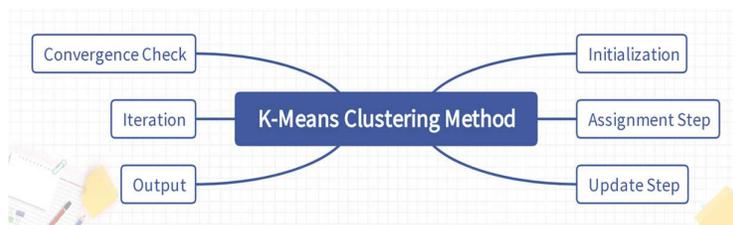
Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. Data Selection: Data perceraian dari Badan Pusat Statistik Jawa Timur tahun 2020-2022 digunakan dengan atribut nama kabupaten/kota dan jumlah kasus perceraian.
2. Preprocessing: Data yang hilang diimputasi dengan rata-rata untuk menjaga integritas data.
3. Transformation: Normalisasi data menggunakan Min-Max Scaling untuk menghindari dominasi variabel tertentu.
4. Clustering: Algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan karakteristik serupa.
5. Interpretation: Hasil clustering dievaluasi untuk mendapatkan wawasan pola distribusi perceraian.

Interpretation / Evaluation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

Algoritma K-means

Metode K-Means Clustering adalah algoritma yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan karakteristik. Algoritma ini bekerja melalui beberapa tahap iteratif untuk menghasilkan pengelompokan data yang optimal. dapat di lihat pada gambar 2.1



gambar 2.1. Alur Metode K-Means

Metode K-Means Clustering adalah algoritma yang digunakan untuk mengelompokkan data ke dalam sejumlah cluster berdasarkan kesamaan karakteristik. Algoritma ini bekerja melalui beberapa tahap iteratif untuk menghasilkan pengelompokan data yang optimal. Tahapan utama dalam algoritma ini meliputi:

1. Inisialisasi
Pada tahap ini, centroid awal ditentukan secara acak. Centroid merupakan representasi pusat dari masing-masing cluster yang akan diperbarui selama proses iterasi.
2. Langkah Penugasan (Assignment Step)
Setiap data dalam dataset ditugaskan ke cluster yang memiliki centroid terdekat. Kedekatan ini diukur menggunakan Euclidean Distance, yang menghitung jarak geometris antara data dan centroid.
3. Langkah Pembaruan (Update Step)
Setelah semua data ditugaskan ke cluster tertentu, centroid diperbarui berdasarkan rata-rata posisi dari seluruh data dalam cluster tersebut. Proses ini bertujuan untuk merepresentasikan pusat baru dari cluster.
4. Pemeriksaan Konvergensi (Convergence Check)
Algoritma akan terus melakukan iterasi hingga posisi centroid stabil, atau hingga mencapai jumlah iterasi maksimum yang telah ditentukan. Stabilitas ini menandakan bahwa tidak ada lagi perubahan signifikan dalam pengelompokan data.
5. Hasil Akhir (Output)
Setelah proses konvergensi tercapai, algoritma menghasilkan cluster final yang merepresentasikan pengelompokan data sesuai dengan karakteristik yang diamati. Posisi centroid untuk setiap cluster juga menjadi bagian dari hasil akhir.

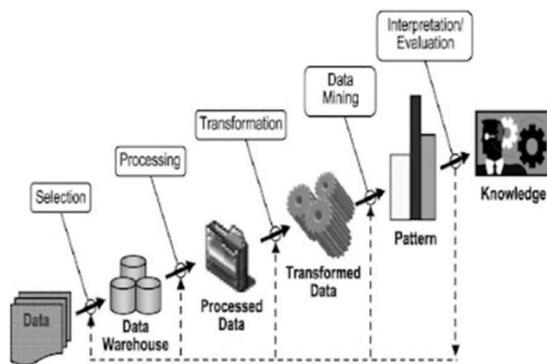
Davies Bouldin Index

Nilai DBI digunakan untuk mengevaluasi kualitas clustering. Semakin kecil nilai DBI, semakin baik pemisahan antar cluster dan kekompakan dalam cluster. Semakin optimal skema pengelompokan yang dihasilkan. Rumus DBI adalah sebagai berikut:

$$DBI = \frac{1}{M} \sum_{j=1}^m \max_{j \neq k} R_{jk}$$

Tahap Perancangan

Knowledge Discovery in Database (KDD) adalah proses untuk mengidentifikasi informasi yang relevan serta pola-pola tersembunyi dalam data. Informasi ini ditemukan dalam basis data berukuran besar yang sebelumnya tidak diketahui dan memiliki potensi untuk memberikan manfaat. Penelitian ini menggunakan teknik knowledge Discovery In Database (KDD) untuk menganalisa data. Teknik terdiri dari beberapa tahap, yaitu :



Gambar 2.2. Tahapan Knowledge of Database

1. Data selection

Seleksi data statistik adalah teknik untuk meminimalkan jumlah data yang digunakan dalam teknik mining sambil tetap mempertahankan data aslinya. Hasil pilihan statistik yang diperoleh dari teknik data mining disimpan dalam file yang terpisah dari database produksi. Pada tahap ini, data diambil dari Dataset Badan Pusat statistik Provinsi Jawa Timur

Tabel 2.1. Data Kasus perceraian di

No.	Nama Kabupaten/Kota	Tahun 2020	Tahun 2021	Tahun 2022
1	Kab. Pacitan	1.135	1190	1192
2	Kab. Ponorogo	727	1921	2020
3	Kab. Trenggalek	1659	1678	1777
4	Kab. Tulungagung	2822	2511	3171
5	Kab. Blitar	4045	3669	4096
6	Kab. Kediri	1131	3711	4162
7	Kab. Malang	6707	6370	8195
8	Kab. Lumajang	2442	3034	3361
9	Kab. Jember	354	5864	6779
...
38	Kota Surabaya	5154	5726	6933

2. Data preprocessing

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus penemuan pengetahuan (knowledge discovery). Proses pembersihan meliputi penghapusan duplikasi data, pemeriksaan data yang tidak konsisten, dan perbaikan kesalahan pada data, seperti kesalahan pengetikan. Selain

itu, dilakukan juga proses pengayaan (enrichment), yaitu memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD.

3. Transformation adalah proses mengubah data yang telah dipilih agar sesuai untuk proses data mining. Proses pengkodean dalam penemuan pengetahuan (knowledge discovery) adalah proses kreatif yang sangat tergantung pada jenis atau pola informasi yang ingin ditemukan dalam basis data
4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma dalam data mining sangat beragam bergantung pada tujuan dan keseluruhan proses penemuan pengetahuan (KDD)

5. Interpretation atau evaluation

interpretasi atau evaluasi adalah hasil dari data mining di analisis untuk memastikan bahwa apakah pola atau Informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

III. HASIL

DATA SELECTION DAN PREPOCESING

Pada tahap ini, data dipilih berdasarkan fitur yang terkait dengan analisis kelompok. Data yang digunakan meliputi jumlah kasus perceraian per tahun untuk masing-masing kabupaten/kota di Provinsi Jawa Timur. Data ini diimpor dari file CSV, dan kemudian kolom yang mencakup nama kabupaten/kota dan jumlah kasus perceraian dari tahun 2020 hingga 2022 dipilih untuk seleksi atribut. Agar proses analisis berikutnya tidak terpengaruh, data yang hilang atau kosong diidentifikasi dan diisi dengan metode imputasi rata-rata.

```
# Load dataset
data = pd.read_csv("data_cerai.csv")
print(data.head())

# Kolom yang akan digunakan
selected_data = data[['Nama Kabupaten/Kota', 'Tahun 2020', 'Tahun 2021',
                    'Tahun 2022']]

# Cek data kosong
print("Jumlah data kosong sebelum imputasi:")
print(selected_data.isnull().sum())

# Isi data kosong hanya pada kolom numerik dengan nilai rata-rata
numeric_columns = selected_data.select_dtypes(include=['int64', 'float64']).columns
selected_data[numeric_columns] = selected_data[numeric_columns].fillna(selected_data[numeric_columns].mean())

# Konfirmasi data kosong setelah imputasi
print("Jumlah data kosong setelah imputasi:")
print(selected_data.isnull().sum())
```

Gambar 3.1. Proses Data Selection dan Preprocessing

```

Nama Kabupaten/Kota Tahun 2020 Tahun 2021 Tahun 2022
0 Kab. Pacitan 1135.0 1190.0 1192.0
1 Kab. Ponorogo 727.0 1921.0 2020.0
2 Kab. Trenggalek 1659.0 1678.0 1777.0
3 Kab. Tulungagung 2822.0 2511.0 3171.0
4 Kab. Blitar 4045.0 3669.0 4096.0
Jumlah data kosong sebelum imputasi:
Nama Kabupaten/Kota 1
Tahun 2020 4
Tahun 2021 4
Tahun 2022 4
dtype: int64
Jumlah data kosong setelah imputasi:
Nama Kabupaten/Kota 1
Tahun 2020 0
Tahun 2021 0
Tahun 2022 0
dtype: int64
    
```

Gambar 3.2. Hasil Proses Data Selection dan Preprocessing

Tranformasi Data

Tranformasi Data dilakukan untuk memastikan data dalam format yang siap diolah oleh algoritma K-Means. Nama kabupaten/kota yang merupakan data nominal dikonversi menjadi numerik menggunakan teknik encoding. Selanjutnya, data di-normalisasi menggunakan metode Min-Max Scaling untuk memastikan semua variabel berada dalam rentang nilai yang sama. Normalisasi ini penting untuk menghindari dominasi salah satu variabel terhadap hasil clustering.

```

# Ubah nama kabupaten menjadi indeks numerik
selected_data['Nama Kabupaten/Kota'] = selected_data['Nama Kabupaten/Kota'].astype('category').cat.codes

# Normalisasi data
scaler = MinMaxScaler()
normalized_data = scaler.fit_transform(selected_data.iloc[:, 1:]) # Kolom tahun saja
    
```

Gambar 3.3. Proses Transformasi Data

	Tahun 2020	Tahun 2021	Tahun 2022
0	0.018313	0.010312	0.007605
1	0.011718	0.018624	0.015751
2	0.026783	0.015861	0.013360
3	0.045581	0.025332	0.027074
4	0.065349	0.038498	0.036175

Gambar 3.4. Hasil Tranformasi Data

Pada tahap ini, algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan karakteristik yang serupa. Untuk menentukan jumlah cluster optimal, evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI). DBI digunakan karena mampu menilai kualitas clustering dengan memperhatikan jarak antar cluster dan kekompakan data dalam cluster. Nilai DBI terendah menunjukkan jumlah cluster optimal. Pada penelitian ini menggunakan cluster dengan nilai K2=2, K3= 3, K5=5, K7=7.

```
dbi_scores = [] # Menyimpan hasil DBI untuk setiap jumlah cluster

# Nilai k yang akan diuji
k_values = [2, 3, 5, 7] # Sesuai permintaan

for k in k_values: # Loop dengan nilai k tertentu
    kmeans = KMeans(n_clusters=k, random_state=42)
    labels = kmeans.fit_predict(normalized_data)
    dbi = davies_bouldin_score(normalized_data, labels)
    dbi_scores.append((k, dbi))

# Tambahkan hasil clustering ke DataFrame
cluster_df = normalized_df.copy()
cluster_df['Cluster'] = labels
```

Gambar 3.5. Proses Clustering K-Means

```
# Tampilkan hasil clustering
print(f"\nJumlah Cluster: {k}")
print(f"Nilai DBI: {dbi:.4f}")
print("\nHasil Clustering:")
for cluster in sorted(cluster_df['Cluster'].unique()):
    print(f"Cluster {cluster}: {len(cluster_df[cluster_df['Cluster'] == cluster])} data")
    print("-" * 40)

# Menentukan jumlah cluster optimal
optimal_cluster = min(dbi_scores, key=lambda x: x[1])
print("\nCluster terbaik berdasarkan DBI terendah:")
print("Jumlah cluster optimal:", optimal_cluster[0])
print("Nilai DBI:", optimal_cluster[1])
```

Gambar 3.6. Menampilkan Hasil Clustering

Hasil Cluster K-Means K=2

Proses clustering dengan menggunakan nilai K = 2 menghasilkan 2 cluster dan memperoleh nilai Davies-Bouldin Index (DBI) sebesar 0.6108. Pada cluster 0, terdapat 6 data, sedangkan pada cluster 1, terdapat 33 data kabupaten."

```
Jumlah Cluster: 2
Nilai DBI: 0.6108

Hasil Clustering:
Cluster 0: 6 data
-----
Cluster 1: 33 data
-----
```

Gambar 3.7. Hasil Cluster K=2

Hasil Cluster K-Means K=3

Proses clustering dengan menggunakan nilai $K = 3$ menghasilkan 3 cluster dan memperoleh nilai Davies-Bouldin Index (DBI) sebesar 0.7767. Pada cluster 0, terdapat 8 data; pada cluster 1, terdapat 28 data; dan pada cluster 2, terdapat 3 data."

```
Jumlah Cluster: 3
Nilai DBI: 0.7768

Hasil Clustering:
Cluster 0: 8 data
-----
Cluster 1: 28 data
-----
Cluster 2: 3 data
-----
```

Gambar 3.7. Hasil Cluster K=3

Hasil Cluster Kmeans K=5

Proses clustering dengan menggunakan nilai $K = 5$ menghasilkan 5 cluster dan menghasilkan nilai Davies-Bouldin Index (DBI) sebesar 0.8146. Pada cluster 0, terdapat 3 data; pada cluster 1, terdapat 17 data; pada cluster 2, terdapat 3 data; pada cluster 3, terdapat 13 data; dan pada cluster 4, terdapat 3 data

```
Jumlah Cluster: 5
Nilai DBI: 0.8146

Hasil Clustering:
Cluster 0: 3 data
-----
Cluster 1: 17 data
-----
Cluster 2: 3 data
-----
Cluster 3: 13 data
-----
Cluster 4: 3 data
-----
```

Gambar 3.7. Hasil Cluster K=5

Hasil Cluster Kmeans K=7

Proses clustering dengan menggunakan nilai $K = 7$ menghasilkan 7 cluster dan menghasilkan nilai Davies-Bouldin Index (DBI) sebesar 0.4601. Pada cluster 0, terdapat 2 data; pada cluster 1, terdapat 12 data; pada cluster 2, terdapat 3 data; pada cluster 3, terdapat 14 data; pada cluster 4, terdapat 3 data; pada cluster 5 terdapat 1 data; pada cluster 6 terdapat 4 data.

```
Jumlah Cluster: 7
Nilai DBI: 0.4601

Hasil Clustering:
Cluster 0: 2 data
-----
Cluster 1: 12 data
-----
Cluster 2: 3 data
-----
Cluster 3: 14 data
-----
Cluster 4: 3 data
-----
Cluster 5: 1 data
-----
Cluster 6: 4 data
-----
```

Gambar 3.7. Hasil Cluster $K=7$

berikut pembagian data cluster dengan nilai $K=7$ terdapat pada tabel dibawah ini.

tabel data cluster $K=7$

Cluster nilai $K= 7$	
Kabupaten/Kota	Cluster
Kab. Blitar	Cluster 0
Kab. Sidoarjo	Cluster 0
Kab. Pacitan	Cluster 1
Kab. Ponorogo	Cluster 1
Kab. Trenggalek	Cluster 1
Kab. Bondowoso	Cluster 1

Kab. Situbondo	Cluster 1
Kab. Madiun	Cluster 1
Kab. Magetan	Cluster 1
Kab. Ngawi	Cluster 1
Kab. Bangkalan	Cluster 1
Kab. Pamekasan	Cluster 1
Kab. Sampang	Cluster 1
Kota Pasuruan	Cluster 1
kab. Malang	Cluster 2
Kab. Banyuwangi	Cluster 2
Kota Surabaya	Cluster 2
Kab. Tulungagung	Cluster 3
Kab. Kediri	Cluster 3
Kab. Lumajang	Cluster 3
Kab. Pasuruan	Cluster 3
Kab. Mojokerto	Cluster 3
Kab. Jombang	Cluster 3
Kab. Nganjuk	Cluster 3
Kab. Tuban	Cluster 3
Kab. Sumenep	Cluster 3
Kota Blitar	Cluster 3
Kota Mojokerto	Cluster 3
Kota Batu	Cluster 3
Kota Kediri	Cluster 4
Kota Probolinggo	Cluster 4

Kota Madiun	Cluster 4
Kab. Jember	Cluster 5
Kab. Probolinggo	Cluster 6
Kab. Bojonegoro	Cluster 6
Kab. Lamongan	Cluster 6
Kab. Gresik	Cluster 6

Dari tabel yang ditampilkan diatas diperoleh bahwa Cluster terbaik menggunakan nilai K=7 dari pengelompokan kasus perceraian dan terdapat 7 cluster yaitu cluster 0 terdapat 2 data; pada cluster 1, terdapat 12 data; pada cluster 2, terdapat 3 data; pada cluster 3, terdapat 14 data; pada cluster 4, terdapat 3 data,; pada cluster 5 terdapat 1 data; pada cluster 6 terdapat 4 data dan menghasilkan davies bouldin index 0,460

```
Cluster terbaik berdasarkan DBI terendah:
Jumlah cluster optimal: 7
Nilai DBI: 0.4601148848056876
```

Gambar 3.7. Hasil Cluster Optimal

Evaluasi dan Visualisasi

Setelah clustering dilakukan, hasilnya dievaluasi dan divisualisasikan. Data di setiap cluster dianalisis untuk memahami pola distribusi perceraian. Visualisasi dalam bentuk scatter plot menunjukkan distribusi data dalam dua dimensi, dengan centroid masing-masing cluster ditampilkan untuk memberikan gambaran lokasi pusat data dalam kelompok tersebut. Hasil ini memberikan wawasan mengenai kategori daerah dengan tingkat perceraian tinggi dan rendah.

```
# 1. Visualisasi hasil clustering pada data normalisasi
def plot_clusters(k_values, normalized_data, dbi_scores):
    plt.figure(figsize=(10, 6))

    for k in k_values:
        kmeans = KMeans(n_clusters=k, random_state=42)
        labels = kmeans.fit_predict(normalized_data)

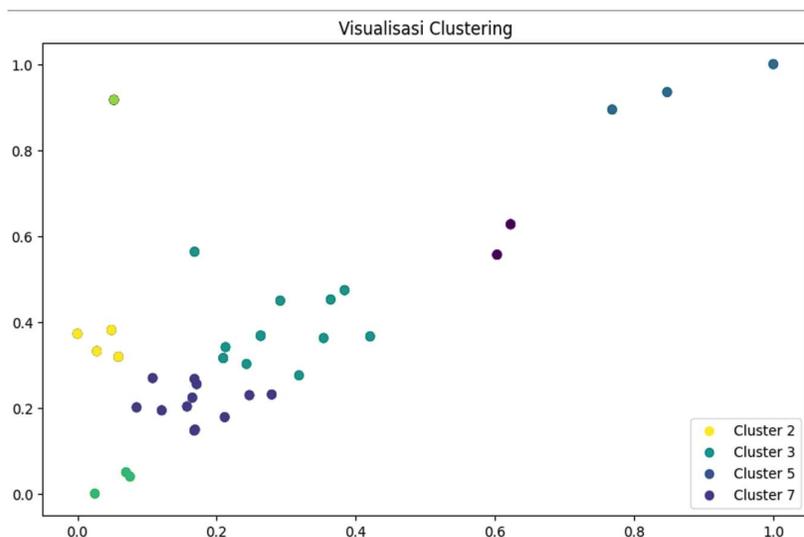
        # Membuat plot dengan warna berbeda untuk setiap cluster
        plt.scatter(normalized_data[:, 0], normalized_data[:, 1], c=labels, cmap='viridis', label=f'Cluster {k}')

    plt.title('Visualisasi Clustering')
    plt.xlabel('Tahun 2020')
    plt.ylabel('Tahun 2021')
    plt.legend(loc='best')
    plt.show()

# 2. Visualisasi grafik DBI untuk memilih jumlah cluster optimal
def plot_dbi_scores(dbi_scores):
    k_vals = [score[0] for score in dbi_scores]
    dbi_vals = [score[1] for score in dbi_scores]

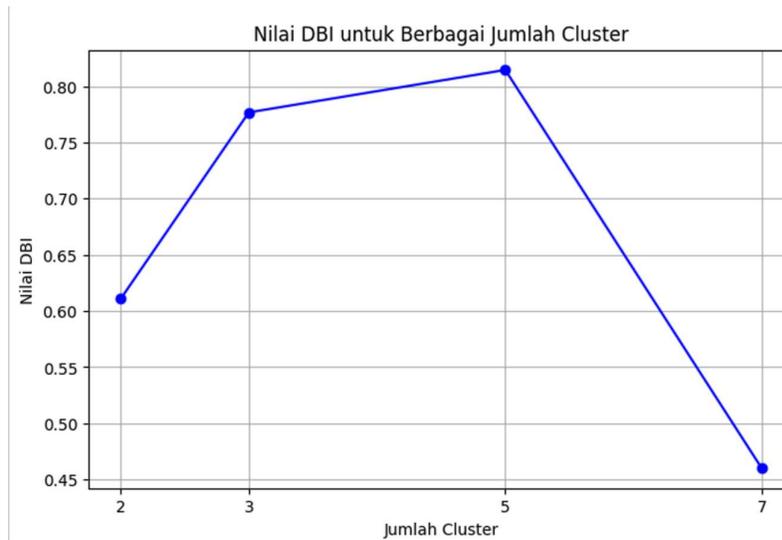
    plt.figure(figsize=(8, 5))
    plt.plot(k_vals, dbi_vals, marker='o', linestyle='-', color='b')
    plt.title('Nilai DBI untuk Berbagai Jumlah Cluster')
    plt.xlabel('Jumlah Cluster')
    plt.ylabel('Nilai DBI')
    plt.xticks(k_vals)
    plt.grid(True)
    plt.show()
```

Gambar 3.7. Proses Evaluasi dan visualisasi



Gambar 3.8. Hasil Visualisasi

Gambar tersebut menunjukkan hasil clustering data berdasarkan dua variabel, yaitu "Tahun 2020" (sumbu X) dan "Tahun 2022" (sumbu Y). Data dikelompokkan menjadi empat cluster yang ditandai dengan warna berbeda: kuning (Cluster 2), hijau-biru (Cluster 3), biru tua (Cluster 5), dan ungu tua (Cluster 7). Visualisasi ini membantu mengidentifikasi pola distribusi data antara kedua tahun, di mana setiap cluster merepresentasikan kelompok dengan karakteristik serupa berdasarkan perubahan nilai antar tahun.



Gambar 3.9 Visualisasi Nilai DBi dalam jumlah cluster

pada gambar diatas menunjukkan nilai Davies-Bouldin Index (DBI) untuk berbagai jumlah cluster. Sumbu X merepresentasikan jumlah cluster (2, 3, 5, dan 7), sedangkan sumbu Y menunjukkan nilai DBI. Semakin rendah nilai DBI, semakin baik kualitas clustering. Dari grafik terlihat bahwa nilai DBI menurun signifikan saat jumlah cluster bertambah dari 5 ke 7, menunjukkan bahwa model clustering dengan 7 cluster memberikan hasil yang lebih optimal dibandingkan jumlah cluster lainnya.

IV. KESIMPULAN

Penelitian ini menganalisis pola data perceraian melalui pendekatan clustering untuk memahami karakteristik distribusi kasus perceraian dari tahun 2020 hingga 2021. Data berhasil dikelompokkan ke dalam beberapa cluster, yang masing-masing merepresentasikan karakteristik berbeda terkait jumlah dan tren perceraian antar tahun. Dari visualisasi clustering, terlihat bahwa kelompok-kelompok tersebut menunjukkan pola peningkatan, penurunan, atau kestabilan jumlah perceraian.

Evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa model clustering dengan 7 cluster memberikan kualitas terbaik dibandingkan dengan jumlah cluster lainnya. Hal ini ditunjukkan oleh nilai DBI yang lebih rendah, yang mengindikasikan pemisahan antar cluster yang baik serta tingkat homogenitas yang tinggi di dalam masing-masing cluster.

Hasil clustering ini memberikan wawasan mendalam terkait faktor-faktor penyebab perceraian di setiap cluster. Temuan ini dapat dimanfaatkan oleh pengambil kebijakan atau instansi terkait untuk merancang program dan kebijakan yang lebih efektif dan tepat sasaran dalam menangani kasus perceraian berdasarkan karakteristik di setiap kelompok.

Penelitian selanjutnya disarankan untuk mempertimbangkan variabel tambahan, seperti faktor ekonomi atau pendidikan, guna memperkaya analisis. Selain itu, pengumpulan data secara berkala perlu dilakukan untuk mengidentifikasi tren perceraian di masa depan dengan lebih baik.

V. DAFTAR PUSTAKA

- [1] Yunus Muhammad, Aziz Abdul, hasanah Nurul, Khasanah Jama'atun, and A'im5, "PENGARUH PERNIKAHAN DINI TERHADAP TINGKAT PERCERAIAN DI KECAMATAN ABUNG BARAT," *Jurnal Multidisiplin Ilmu*, vol. Vol. 2 (1), pp. 35–44, Jan. 2023, doi: <https://doi.org/10.31004/koloni.v2i1.381>.

- [2] J. Jukim ; | P-Dhani Nadiatusholikha ; Putri, S. S. Syarifa, and N. . Rofiq, “Lanjar Indah Kusumawardhany, Rizkyta Dwi Fatimah, Nailus Syarifa, Salsabila Swastika Putri,” *Jurnal Ilmiah Multidisiplin*, vol. 3, no. 4, pp. 95–102, 2024, doi: 10.56127/jukim.v3i04.
- [3] Winarsih, Junadi, and Taufiq Ahmad, “EFEKTIVITAS PROGRAM BIMBINGAN PERKWINAN KEMENTERIAN AGAMA DALAM MENCEGAH PERCERIAN DI KABUPATEN BOJONEGORO TAHUN 2018-2020,” *JIAN - Jurnal Ilmiah Administrasi negara*, vol. Vol. 6 (2), pp. 13–19, 2022, doi: <https://doi.org/10.56071/jian.v6i2.652>.
- [4] E. Purwaningsih and E. Nurelasari, “Implementasi Metode K-Means Clustering Dengan Davies Bouldin Index Pada Analisis Faktor Penyebab Perceraian,” *INFORMATION MANAGEMENT FOR EDUCATORS AND PROFESSIONALS*, vol. 7, no. 2, pp. 134–143, 2023, [Online]. Available: <https://databoks.katadata.co.id/>
- [5] N. A. Sudibyoy, A. Iswardani, K. Sari, S. Suprihatiningsih, U. Duta, and B. Surakarta, “PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI INDONESIA,” *Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, vol. 1, no. 3, pp. 199–207, doi: 10.46306/lb.v1i3.
- [6] L. Magdalena and R. Fahrudin, “Penerapan Data Mining Untuk Koperasi Se-Jawa Barat Menggunakan Metode Clustering pada Kementerian Koperasi dan UKM,” 2019. doi: <https://doi.org/10.51920/jd.v9i2.120>.
- [7] I. Nozomi, “PENERAPAN DATA MINING UNTUK PERINGATAN DINI BANJIR MENGGUNAKAN METODE KLASSTERING K-MEANS (STUDI KASUS KOTA PADANG),” Bulan Juni, 2023.
- [8] D. A. Fakhri, S. Defit, and Sumijan, “Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering,” *Jurnal Informasi dan Teknologi*, pp. 160–166, Sep. 2021, doi: 10.37034/jidt.v3i3.137.
- [9] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa,” 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [10] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan.
- [11] M. Triandini, S. Defit, and G. W. Nurcahyo, “Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia,” *Jurnal Informasi dan Teknologi*, pp. 167–173, Sep. 2021, doi: 10.37034/jidt.v3i3.120.
- [12] P. Apriyani, A. R. Dikananda, and I. Ali, “Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi,” *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, Mar. 2023, doi: 10.56211/helloworld.v2i1.230.
- [13] Anggreini Novita Lestari, “TEKNIK CLUSTERING DENGAN ALGORITMA K-MEDOIDS UNTUK MENANGANI STRATEGI PROMOSI DI POLITEKNIK TEDC BANDUNG,” Dec. 2019. [Online]. Available: <http://tip.ppj.unp.ac.id>
- [14] Elfaladonna Febie and RahmadaniAyu, “ANALISA METODE CLASSIFICATION-DECISSION TREE DAN ALGORITMA C.45 UNTUK MEMPREDIKSI PENYAKIT DIABETES DENGAN MENGGUNAKAN APLIKASI RAPID MINER,” *Science and Information Technology*, vol. 2, pp. 10–7, Apr. 2019, doi: <https://doi.org/10.31598/sintechjournal.v2i1.293>.
- [15] Fadlina, “DATA MINING UNTUK PREDIKSI PENYAKIT BATU EMPEDU DENGAN ALGORITMA C45 APLIKASI RAPID MINER,” Jul. 2023. doi: <https://doi.org/10.58369/git.v1i2.126>.
- [16] C. Zai and T. Komputer, “IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA.”
- [17] E. Tachi and N. & Andri, “Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN,” 2021.
- [18] N. A. Sudibyoy, A. Iswardani, K. Sari, S. Suprihatiningsih, U. Duta, and B. Surakarta, “PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI INDONESIA,” *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 1, no. 3, pp. 199–207, Dec. 2020, doi: 10.46306/lb.v1i3.
- [19] I. Parlina, A. Perdana Windarto, A. Wanto, Mr. Lubis, D. Amik Tunas Bangsa Pematangsiantar, and D. A. STIKOM Tunas Bangsa JIjendral Sudirman Blok No, “MEMANFAATKAN ALGORITMA K-MEANS DALAM

MENENTUKAN PEGAWAI YANG LAYAK MENGIKUTI ASESSMENT CENTER UNTUK CLUSTERING PROGRAM SDP,” Jan. 2018.