



Segmentasi Faktor Perceraian berdasarkan Provinsi di Indonesia Tahun 2024 dengan K-Means dan DBSCAN

Selly Rizkiyah ^{a,1}, Indira Zein Rizqin ^{b,2,*}, Milla Akbarany Bakhtiar Putri ^{b,3}, Shindi Shella May Wara ^{b,4}, Wahyu Syaifulullah Jauharis Saputra ^{b,5}

^{a,b} Universitas Pembangunan Nasional "Veteran" Jawa Timur, Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294, Indonesia
¹ 23083010010@student.upnjatim.ac.id; ² 23083010015@student.upnjatim.ac.id*; ³ 23083010021@student.upnjatim.ac.id;
⁴ shindi.shella.fasilkom@upnjatim.ac.id; ⁵ wahyu.s.j.saputra.if@upnjatim.ac.id

* corresponding author

ABSTRACT

ARTICLE INFO

Perceraian merupakan fenomena sosial yang kompleks dan terus mengalami peningkatan di Indonesia. Berdasarkan data dari 34 provinsi, perceraian dipengaruhi oleh berbagai faktor, baik internal maupun eksternal rumah tangga. Penelitian ini bertujuan membandingkan metode clustering terbaik antara *K-Means* dan *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* dalam menggambarkan faktor-faktor utama penyebab perceraian berdasarkan data nasional serta meninjau literatur yang relevan menggunakan metode machine learning, khususnya teknik unsupervised learning berupa clustering. Faktor dominan yang ditemukan antara lain adalah perselisihan dan pertengkaran terus-menerus, permasalahan ekonomi, kekerasan dalam rumah tangga, tindakan meninggalkan salah satu pihak, dan perselingkuhan. Penelitian ini menggunakan algoritma *K-Means* dan *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* untuk membandingkan hasilnya. Diketahui pemodelan terbaik dengan perbandingan Silhouette Score adalah *DBSCAN* sebesar 0.331. *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* dengan cluster optimal didapat dari kombinasi parameter epsilon 2.9 dan minimal sampel 2. Hasil clustering kemudian dianalisis lebih lanjut untuk mengevaluasi distribusi data dan mengidentifikasi karakteristik dominan di tiap kluster. Temuan ini menunjukkan perlunya pendekatan multidisipliner dalam memahami dan menangani permasalahan perceraian di Indonesia agar dapat menekan angka perceraian dan meningkatkan kualitas kehidupan keluarga.

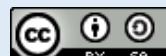
Article history

Received: 16 Juni 2025
Revised: 30 Juni 2025
Accepted: 10 Juli 2025

Keywords

1. Perceraian
2. Faktor Penyebab
3. Clustering
4. Provinsi
5. DBSCAN

This is an open access article under the [CC-BY-SA](#) license.



I. Pendahuluan

Berdasarkan data Badan Pusat Statistik (BPS) [1], jumlah perkara perceraian di Indonesia mencapai 394.608 pada tahun 2024. Alasannya beragam mulai dari masalah ekonomi hingga ketidakcocokan dalam rumah tangga. Fenomena ini menunjukkan bahwa perceraian bukan lagi kasus yang bersifat kasuistik, melainkan persoalan sosial yang terus meningkat dan membutuhkan perhatian khusus.

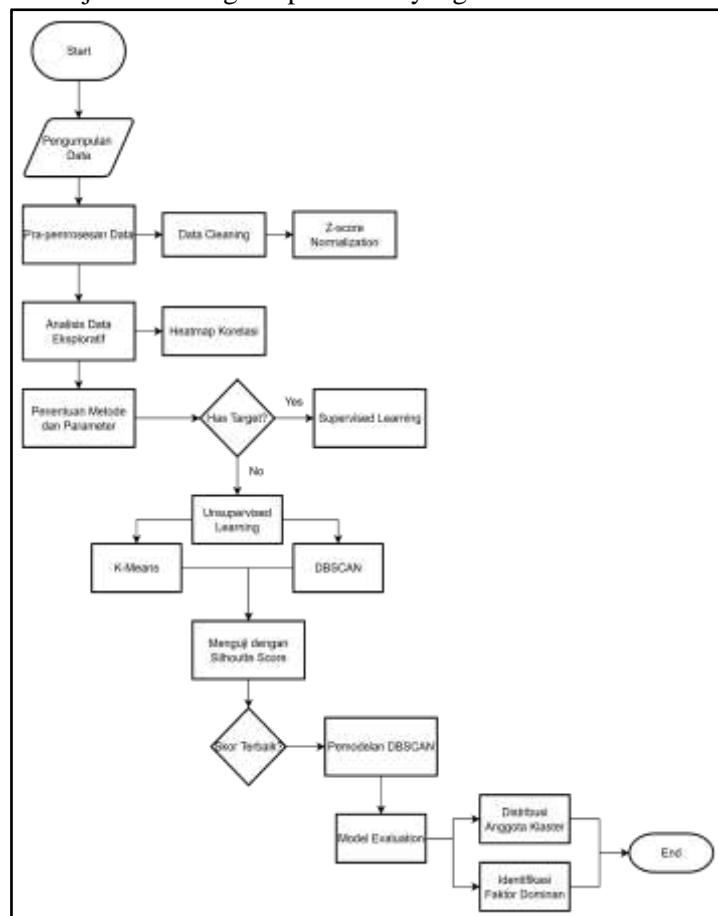
Perkawinan merupakan ikatan lahir batin antara seorang pria dan seorang wanita sebagai suami istri dengan tujuan membentuk keluarga atau rumah tangga yang bahagia dan kekal berdasarkan Ketuhanan Yang Maha Esa [2]. Keterikatan dalam keluarga melalui pernikahan yang menggabungkan pria dan wanita dengan aspirasi serta tujuan untuk mencapai kebahagiaan bersama. Namun, hal ini tidak berlaku global pada tiap pernikahan, banyak pasangan yang menghadapi masalah-masalah dalam hubungan pernikahan yang dapat berujung pada retaknya status rumah tangga mereka [3]. Meskipun demikian, Perceraian sendiri dapat terjadi melalui beberapa mekanisme, yaitu cerai talak (permohonan cerai oleh suami), cerai gugat (permohonan cerai oleh istri), dan berakhirnya perkawinan akibat kematian salah satu pasangan [4].

Indonesia mencatat bahwa penyebab perceraian sangat beragam dan terus menunjukkan angka signifikan. Sebanyak 13 faktor utama penyebab perceraian yakni permasalahan ekonomi, pertengkarannya terus menerus, dan lain sebagainya [1]. Selain itu, tekanan sosial, kegagalan komunikasi, perbedaan visi hidup, dan kurangnya kesiapan dalam menjalani rumah tangga sangat berpengaruh terhadap retaknya hubungan pernikahan [5].

Penelitian ini bertujuan membandingkan metode clustering terbaik antara *K-Means* dan *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) dalam menggambarkan faktor-faktor utama penyebab perceraian berdasarkan data nasional serta meninjau literatur yang relevan menggunakan metode machine learning, khususnya teknik *unsupervised learning* berupa *clustering*. Pada penelitian sebelumnya disebutkan bahwa metode *K-Means* cocok untuk data yang homogen dengan menunjukkan segmentasi yang kuat dan terdistribusi secara merata [6]. Sedangkan metode *DBSCAN* lebih sesuai digunakan untuk mendeteksi klaster dengan bentuk yang tidak beraturan dan dapat menangani outlier secara efektif. Selain itu, hasil penelitian terdahulu menunjukkan bahwa *DBSCAN* menghasilkan klaster dengan nilai indeks validitas menggunakan algoritma *Silhouette* [7].

2. Metode dan Penjelasan

Penelitian kami memiliki urutan alur pada Gambar 1. Flowchart. Berikut ini adalah gambar urutan alur penelitian untuk menjelaskan langkah penelitian yang dilakukan.



Gambar 1. Flowchart

2.1 Data dan Sumber Data

Dataset yang digunakan adalah data sekunder yang diperoleh dari laman resmi BPS dengan judul “Jumlah Perceraian Menurut Provinsi dan Faktor Penyebab Perceraian (perkara), 2024” yang bersumber dari Mahkamah Agung (Dirjen Badan Peradilan Agama) [5]. Dataset ini berisi 38 baris yang memuat provinsi di Indonesia dan 15 kolom yang memuat faktor penyebab perceraian, seperti Faktor Perceraian - Zina, Faktor Perceraian - Mabuk, dan lain-lain. Data ini memiliki cukup variasi dalam hal faktor penyebab perceraian, sehingga memungkinkan pemrosesan menggunakan *clustering*.

2.2 Teknik dan Analisis Data

2.2.1 Pra Pemprosesan Data

Sebelum dilakukan proses *clustering*, data terlebih dahulu diperiksa dan dibersihkan (*data cleaning*) dari nilai kosong atau tidak valid menggunakan rasio. Selanjutnya, data dinormalisasi untuk menyamakan skala antar variabel menggunakan *Z-score normalization*, agar tidak terjadi dominasi nilai dari variabel tertentu.

2.2.2 Analisis Data Eksploratif

Analisis data eksploratif atau yang biasa disebut *EDA* merupakan langkah penting sebelum pemodelan klasifikasi, karena bertujuan untuk memahami karakteristik dan struktur data secara menyeluruh. Salah satu metode yang digunakan dalam *EDA* adalah visualisasi korelasi antar variabel menggunakan *heatmap*. Sebelum dilakukan visualisasi tentunya dilakukan tahap sebelumnya yaitu pra pemrosesan data mencakup *cleaning*, *encoding*, standarisasi [8]. *Heatmap* korelasi kemudian digunakan untuk memvisualisasikan hasil matriks korelasi tersebut. Warna pada *heatmap* menunjukkan kekuatan dan arah hubungan antar variabel [9].

2.2.3 Menentukan Jumlah Cluster Terbaik

Penentuan jumlah cluster optimal dalam penelitian ini dilakukan melalui perbandingan kinerja algoritma *K-Means* dan *DBSCAN* (*Density-Based Spatial Clustering of Applications with Noise*) menggunakan *Silhouette Score* sebagai indikator utama. Kedua algoritma diimplementasikan pada dataset yang sama, kemudian dievaluasi berdasarkan kemampuan membentuk *cluster* yang kompak dan terpisah dengan baik. Hasil perhitungan *Silhouette Score* menunjukkan performa *DBSCAN* lebih unggul dibanding *K-Means* pada kasus ini, sehingga penentuan parameter optimal difokuskan pada *epsilon* (ϵ) dan minimum sample (minPts) untuk *DBSCAN* [10].

DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data [7]. Tidak seperti *K-Means* yang memerlukan jumlah *cluster* di awal, *DBSCAN* mengelompokkan data berdasarkan kepadatan (*density*) dan membutuhkan dua parameter utama, yaitu:

- *eps*: jarak maksimum antar titik yang masih dianggap dalam satu wilayah *cluster*
- *min_samples*: jumlah minimum tetangga dalam radius *eps* untuk membentuk sebuah *cluster*

Untuk menentukan nilai *eps* yang optimal, dilakukan eksperimen dengan beberapa nilai *eps*, kemudian masing-masing hasil *cluster* dievaluasi menggunakan *Silhouette Score*. Setelah nilai *eps* terbaik diperoleh (misalnya *eps* = 2.9 dan *min_samples* = 2), model dibentuk menggunakan *library* scikit-learn.

Validasi visual menggunakan *Yellow Brick* memperkuat keputusan ini dengan menampilkan letak *cluster* dengan jelas sehingga memudahkan interpretasi kualitas *cluster* yang dihasilkan. Pendekatan ini sejalan dengan penelitian lokal yang menunjukkan bahwa *DBSCAN* dengan parameter yang tepat mampu menghasilkan *clustering* yang efektif dan memiliki nilai *Silhouette coefficient* tinggi.

2.2.4 Pemodelan dengan DBSCAN

Pada tahap pemodelan klasifikasi, data dimodelkan menggunakan algoritma *DBSCAN* yang diimporkan dari pustaka scikit-learn. *DBSCAN* dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kepadatan, serta kemampuannya mendeteksi *noise* atau *outlier* tanpa perlu menentukan jumlah *cluster* secara eksplisit. Parameter utama yang dioptimasi adalah *epsilon* (*eps*),

yaitu jarak maksimum antara dua titik agar dianggap berada dalam satu cluster, dan *min_samples*, yaitu jumlah minimum titik dalam radius *eps* untuk membentuk sebuah core point [11].

2.2.5 Evaluasi Hasil

Setelah mengetahui metode terbaik yaitu *DBSCAN*, dilakukan pemodelan *clustering* menggunakan algoritma *DBSCAN*. Tahap selanjutnya adalah evaluasi hasil klasterisasi untuk memahami karakteristik masing-masing cluster dan melihat sejauh mana metode *clustering* berhasil mengelompokkan provinsi dengan pola penyebab perceraian yang serupa. Evaluasi ini dilakukan melalui beberapa langkah berikut:

- Distribusi Jumlah Data per *cluster*

Dilakukan penghitungan jumlah provinsi dalam setiap *cluster* untuk mengetahui distribusi anggota *cluster*. Tujuannya adalah melihat keseimbangan jumlah data dalam masing-masing *cluster*.

- Identifikasi Faktor Dominan di Tiap *cluster*

Dari hasil perhitungan rata-rata, kemudian diambil 5 faktor teratas (*Top-5*) pada setiap *cluster*. Langkah ini bertujuan untuk merumuskan profil penyebab perceraian paling signifikan di setiap kelompok provinsi.

3. Hasil dan Pembahasan

Data mentah 34 provinsi yang diambil dari BPS Indonesia pasti diperlukan pra-proses sebelum memasuki tahap analisis dan pemodelan. Untuk itu, diperlukan tahap pra-proses seperti menangani *missing values*, melakukan transformasi data, standarisasi hingga ke pemodelan serta evaluasi.

3.1 Menangani *missing values*

Sebelum dilakukan tahap *pre-processing*, perlu dilakukan pemeriksaan terhadap nilai *missing values*, pemeriksaan nilai null mencakup seperti berikut:

Tabel 1. Missing values

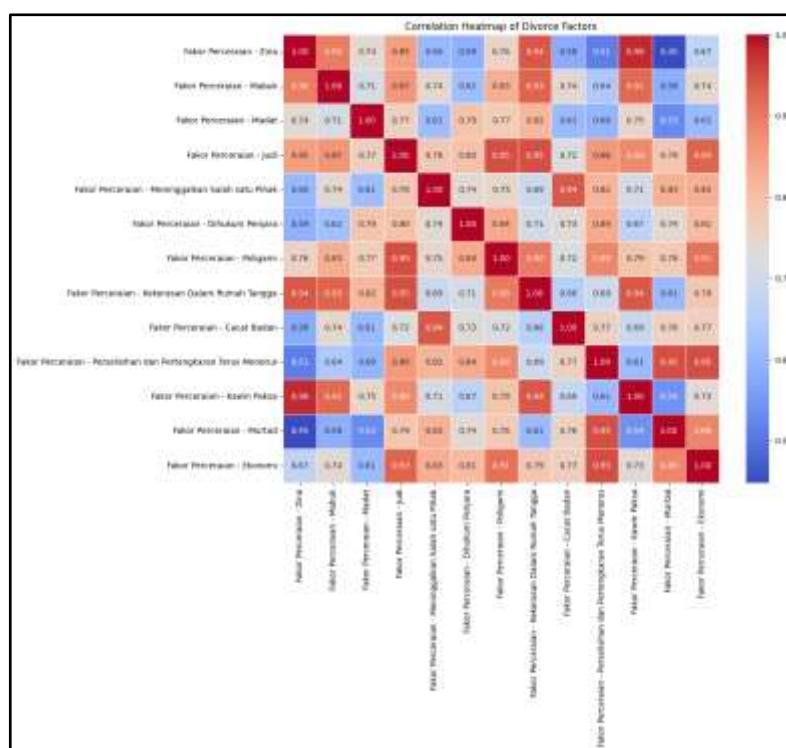
Faktor Penyebab Perceraian	Jumlah Missing Values
Zina	2
Mabuk	1
Madat (penggunaan narkoba)	10
Judi	0
Meninggalkan salah satu pihak	0
Dihukum penjara	0
Poligami	3
Kekerasan Dalam Rumah Tangga (KDRT)	0
Cacat badan	5
Perselisihan dan pertengkarannya terus-menerus	0
Kawin paksa	13
Murtad (keluar dari agama)	1
Ekonomi	0
Jumlah Total Kasus	35

Berdasarkan Tabel 1. Missing values, terdapat sejumlah nilai yang hilang (*missing values*) pada hampir semua kolom faktor perceraian, dengan variasi jumlah yang berbeda-beda. Penanganan nilai yang hilang pada penelitian ini dilakukan dengan pendekatan berbasis rasio kontribusi faktor terhadap total perceraian. Proses ini diawali dengan mengisi nilai kosong atau nol pada kolom "Faktor Perceraian - Jumlah" menggunakan penjumlahan dari seluruh faktor penyebab perceraian yang tersedia pada baris tersebut. Kemudian, membuat kolom rasio untuk setiap faktor. Selanjutnya,

dilakukan perhitungan rasio rata-rata nasional dari masing-masing faktor terhadap total perceraian, berdasarkan data provinsi yang lengkap. Nilai rata-rata rasio tersebut kemudian digunakan untuk mengestimasi nilai hilang pada masing-masing faktor, yaitu dengan mengalikan rasio rata-rata dengan total perceraian di provinsi terkait. Pendekatan ini dinilai lebih representatif dibandingkan metode imputasi sederhana seperti mean atau median, karena tetap mempertimbangkan proporsi faktor terhadap total perceraian serta mempertahankan karakteristik tiap provinsi. Dengan demikian, data yang dihasilkan tetap mencerminkan pola distribusi yang realistik dalam konteks regional.

3.2 Analisis Data Eksploratif

Setelah data telah dilakukan pembersihan, selanjutnya melalui *Heatmap* pada Gambar 1 diketahui korelasi antar variabel. *Heatmap* dibawah ini menunjukkan hubungan antar faktor penyebab perceraian di Indonesia. Warna yang semakin mendekati merah menandakan korelasi positif yang kuat, sedangkan warna biru menunjukkan korelasi yang lebih lemah. Terlihat bahwa sebagian besar faktor memiliki korelasi yang cukup tinggi satu sama lain, terutama faktor-faktor seperti “Zina”, “Kekerasan Dalam Rumah Tangga”, “Poligami”, dan “Kawin Paksa” yang menunjukkan nilai korelasi di atas 0.9 antar sesamanya. Hal ini mengindikasikan bahwa provinsi yang memiliki angka tinggi pada salah satu faktor tersebut cenderung juga memiliki angka tinggi pada faktor lainnya.



Gambar 2. Heatmap correlation

Dari Gambar 2. Heatmap correlation, faktor “Perselisihan dan Pertengkaran Terus Menerus” juga menunjukkan korelasi yang cukup tinggi dengan beberapa faktor seperti “Judi” dan “Murtad”, dan “ekonomi”. Di sisi lain, terdapat beberapa faktor seperti “Murtad” dan “Zina” yang memiliki korelasi lebih rendah terhadap faktor-faktor lain, meskipun masih berada dalam rentang moderat (sekitar 0.5–0.7). Hal ini menunjukkan bahwa meskipun ada hubungan, faktor-faktor tersebut mungkin lebih berdiri sendiri dibanding faktor lainnya. Secara keseluruhan, *heatmap* ini memberikan gambaran bahwa banyak faktor perceraian saling berkaitan erat, yang dapat mempengaruhi hasil clustering karena kesamaan pola antar provinsi

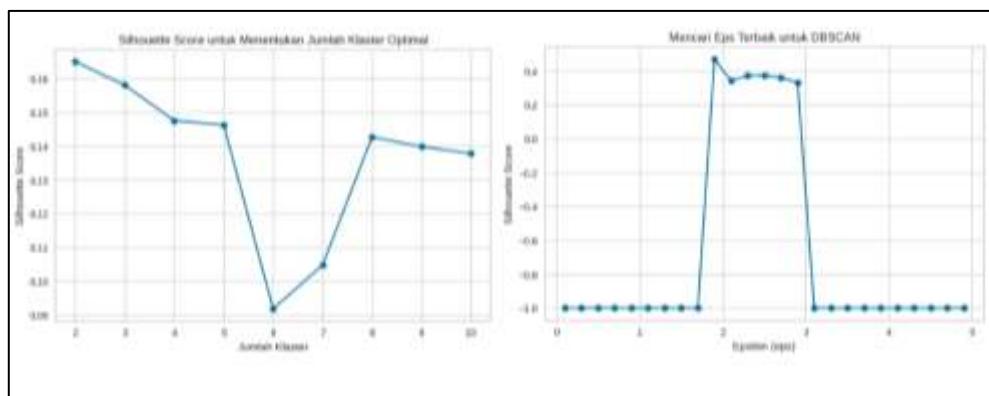
3.3 Penentuan Cluster Terbaik

Setelah menyelesaikan tahapan pembersihan data dan analisis data eksploratif untuk mengetahui karakteristik data, langkah selanjutnya adalah penentuan jumlah *cluster* terbaik sebelum dilakukan

pemodelan lebih lanjut pada kedua algoritma yang diuji, yaitu *K-Means* dan *DBSCAN*. Kedua metode ini memiliki pendekatan berbeda dalam menentukan jumlah cluster dan karakteristik pengelompokan data.

1. Perbandingan *Silhouette Score* *K-Means* dan *DBSCAN*

Pada *K-Means*, skor ini mengukur seberapa mirip suatu data dengan klasternya sendiri dibandingkan dengan *cluster* lainnya. Skor ini berada pada rentang -1 hingga 1, dengan nilai mendekati 1 menunjukkan pengelompokan yang baik. Sementara, pada *DBSCAN* dilakukan dengan pencarian jumlah *cluster* terbaik dengan nilai epsilon dan minimal sampel. *DBSCAN* yang merupakan metode berbasis kepadatan, menawarkan keuntungan tambahan dalam hal kemampuan untuk mengidentifikasi *cluster* dengan bentuk yang tidak teratur dan untuk menangani outlier dengan lebih efektif [6]. *Silhouette Score* juga dapat diamati dengan grafik, menghasilkan hasil seperti berikut ini:

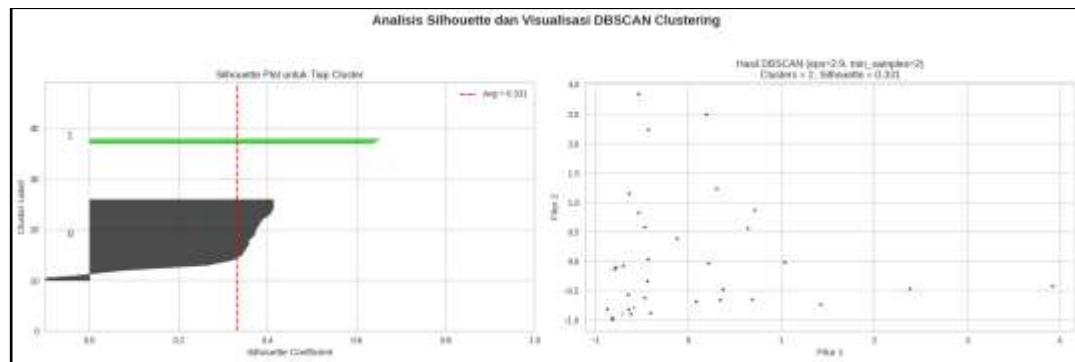


Gambar 3. Visualisasi *Silhouette Score* dengan *K-Means* (kiri) dan *DBSCAN* (kanan)

Dari grafik Gambar 3. Visualisasi *Silhouette Score* dengan *K-Means* (kiri) dan *DBSCAN* (kanan) Silhouette score oleh *K-Means* yaitu 0.164 pada jumlah *cluster* terbaik 2. Sedangkan untuk *DBSCAN* diperoleh *silhouette score* 0.331 saat menggunakan nilai parameter 2.9. Nilai *eps* sebesar 2,9 ditentukan berdasarkan hasil eksplorasi nilai *eps* menggunakan *k-distance graph*, di mana titik siku (*elbow point*) berada pada kisaran nilai tersebut. Sementara itu, nilai minimal sampelnya adalah 2. dipilih berdasarkan karakteristik dataset yang memiliki kepadatan rendah sehingga membutuhkan nilai minimum sampel yang kecil agar dapat membentuk klaster yang relevan. Dengan menentukan jumlah kelompok sama yaitu 2, metode *DBSCAN* menunjukkan performa yang lebih baik dibandingkan *K-Means*. Ini menunjukkan bahwa hasil klaster yang dibentuk memiliki pemisahan antar-klaster yang lebih jelas dan kohesi internal yang lebih baik. Oleh karena itu, metode *DBSCAN* dipilih untuk digunakan dalam proses pemodelan dan analisis klaster selanjutnya.

2. Validasi dengan Visualisasi *Yellowbrick*

Selain menggunakan grafik dan skor, dilakukan juga validasi dengan visualisasi menggunakan *Yellowbrick*, untuk melihat apakah pengelompokan *cluster* dengan *K-Means* memiliki kesalahan berbentuk negatif atau tidak dan untuk melihat sebaran nilai *Silhouette* pada tiap *cluster*. Hasilnya seperti di bawah ini:



Gambar 4. Visualisasi Yellowbrick untuk jumlah cluster 2

Dari Gambar 4. Visualisasi Yellowbrick untuk jumlah cluster 2 (kiri) saat mencoba dengan ϵ 2.9 dan minimal sampel 2, menunjukkan distribusi yang cukup merata, meskipun terdapat sebagian nilai *Silhouette* negatif yang menunjukkan beberapa data berada dekat perbatasan antar *cluster*. Namun, ini dapat diabaikan karena nilainya yang cenderung sangat sedikit. Sementara itu, pada gambar 3 (kanan) menunjukkan jika sebaran cluster terlihat acak. Dapat disimpulkan jika jumlah *cluster* optimal dengan *DBSCAN* menghasilkan jumlah sebanyak 2 *cluster* terbaik.

3.4 Pemodelan dan Hasil Clustering Menggunakan DBSCAN

Setelah menentukan model terbaik yaitu dengan *DBSCAN*, dengan jumlah penentuan parameter terbaik, dilakukan pemodelan menggunakan algoritma *DBSCAN* pada data. Dengan menggunakan parameter *epsilon* = 2.9, sehingga akan mengurangi *noise* dan dapat melakukan klasterisasi dengan optimal, ditambah juga *min_samples* = 2. *Min_samples* menentukan jumlah minimum titik yang diperlukan untuk membentuk suatu cluster [12]. Berikut ini adalah hasil *clustering* provinsi dengan pemodelan *DBSCAN*.

Tabel 2. Provinsi Tiap Cluster dengan DBSCAN

Cluster	Provinsi
-1	Bali DI Yogyakarta Gorontalo Jambi Jawa Timur Kalimantan Barat Kalimantan Timur Kepulauan Bangka Belitung Maluku Utara Nusa Tenggara Timur Papua Papua Barat Sulawesi Barat Sulawesi Tengah Sulawesi Utara
0	Aceh Banten Bengkulu DKI Jakarta Kalimantan Selatan Kalimantan Tengah Kalimantan Utara Kepulauan Riau Lampung Maluku

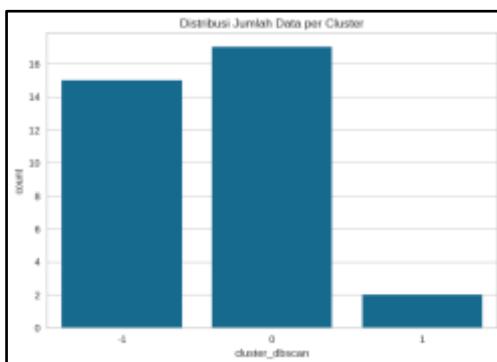
Cluster	Provinsi
	Nusa Tenggara Barat Riau Sulawesi Selatan Sulawesi Tenggara Sumatera Barat Sumatera Selatan Sumatera Utara
1	Jawa Barat Jawa Tengah

Tabel 2. Provinsi Tiap *Cluster* dengan *DBSCAN* menunjukkan pembagian *cluster* tiap provinsi. Hasil dari pemodelan *DBSCAN* juga menghasilkan dua cluster utama, tetapi dengan beberapa provinsi dikategorikan sebagai *noise* (diberi label -1). Pada cluster 0 berisi provinsi-provinsi yang tersebar di berbagai pulau seperti Sumatera, Kalimantan, Sulawesi, dan sebagian di Jawa dan Maluku, namun tidak menunjukkan pola geografis yang jelas. Sementara *cluster* 1 hanya berisi provinsi Jawa Barat dan Jawa Tengah, yang bisa mengindikasikan karakteristik sosial-ekonomi atau budaya yang mirip dan berbeda signifikan dibanding provinsi lain.

Sementara itu, pada *cluster* -1 atau data *noise*, juga terlihat provinsi-provinsi yang acak. Misalnya, Jawa Timur terdeteksi sebagai *noise*, yang artinya data provinsi tersebut tidak cukup dekat dengan titik-titik lain (dalam jarak *epsilon*) untuk tergabung ke dalam salah satu *cluster*. Ini kemungkinan karena provinsi-provinsi dalam *cluster* -1 memiliki nilai fitur yang sangat berbeda dari pola umum, sehingga tidak termasuk ke dalam *cluster* manapun. Didukung data antar provinsi di Indonesia sangat heterogen, baik dari segi geografi, sosial, ekonomi, budaya, maupun demografi. *DBSCAN* dapat menangkap heterogenitas yang tidak bisa ditangani oleh metode seperti *K-Means*.

3.5 Evaluasi Model DBSCAN

Setelah dilakukan tahap pemodelan, selanjutnya dilakukan evaluasi untuk pemahaman lebih mendalam terkait hasil *cluster* yang ada.



Gambar 5. Visualisasi Distribusi Hasil Cluster dengan *DBSCAN*

Berdasarkan Gambar 5. Visualisasi Distribusi Hasil Cluster dengan *DBSCAN*, hasil bar chart distribusi jumlah *cluster*, dapat dilihat jika distribusi data antara *cluster* 0 dan *cluster* 1 sangat timpang dan tidak seimbang, *cluster* 0 memiliki lebih banyak provinsi, dibanding *cluster* 1. Dan terlihat jika masih banyak *noise* pada hasil *cluster*, yang ditandai banyaknya data pada cluster -1. Ini dapat disebabkan karena parameter *eps* yang digunakan terlalu kecil atau *min_samples* terlalu besar, sehingga banyak data tidak masuk ke dalam *cluster* mana pun dan dianggap sebagai *outlier* oleh *DBSCAN*. Meskipun jumlah data yang terdeteksi sebagai *noise* oleh *DBSCAN* lebih besar dibandingkan jumlah anggota klaster, data dengan label *noise* tetap dipertahankan dalam proses pemodelan. Hal ini karena *DBSCAN* secara konseptual mengidentifikasi *noise* sebagai observasi yang secara signifikan berbeda dari pola mayoritas. Titik atau nilai yang tidak termasuk dalam cluster

mana pun diberi label sebagai *noise*, yang dapat sangat bermanfaat saat bekerja dengan data dunia nyata yang tidak teratur [13].

Selanjutnya, maka dilakukan analisis mendalam terkait rata-rata faktor percerai 5 tertinggi di tiap cluster. untuk mengelompokkan wilayah-wilayah di Indonesia berdasarkan faktor-faktor penyebab perceraian. Tujuannya adalah untuk mengidentifikasi pola sosial yang tersembunyi yang bisa menjadi dasar rekomendasi kebijakan. Hasilnya seperti berikut:

Tabel 3. Top 5 Faktor Rata-Rata Perceraian di Tiap Cluster metode DBSCAN

cluster -1 (noise)	cluster 0	cluster 1
Perselisihan dan Pertengkarannya Terus Menerus: 4158.60	Perselisihan dan Pertengkarannya Terus Menerus: 6047.76	Perselisihan dan Pertengkarannya Terus Menerus: 42967.00
Ekonomi: 2353.00	Meninggalkan Salah satu Pihak: 659.94	Ekonomi: 27548.00
Meninggalkan Salah satu Pihak: 683.07	Ekonomi: 576.88	Meninggalkan Salah satu Pihak: 4900.00
Kekerasan Dalam Rumah Tangga: 224.00	Kekerasan Dalam Rumah Tangga: 170.24	Kekerasan Dalam Rumah Tangga: 494.50
Judi: 79.60	Judi: 55.41	Judi: 376.50

Tabel 3. Top 5 Faktor Rata-Rata Perceraian di Tiap *Cluster* metode *DBSCAN* menyajikan rata-rata jumlah kasus dari setiap faktor penyebab perceraian di tiap *cluster*. Tabel ini mempermudah pemahaman karakteristik tiap *cluster*. Misalnya, *cluster -1 (noise)* berisi data yang tidak homogen, namun umumnya masih didominasi faktor perselisihan, ekonomi, dan meninggalkan pasangan. *Cluster* ini mencerminkan data yang dianggap *outlier* oleh model, namun tetap memiliki pola yang bermakna secara sosial. Hal ini menunjukkan bahwa meskipun data dalam *cluster* ini tersebar secara spasial atau statistik, faktor-faktor penyebab perceraian utama tetap konsisten. *Cluster* ini mewakili wilayah dengan dinamika sosial kompleks yang sulit diklasifikasikan secara linier.

Cluster 0 menunjukkan bahwa perselisihan tetap menjadi penyebab utama, namun meninggalkan salah satu pihak menjadi faktor kedua, diikuti oleh ekonomi, kekerasan dalam rumah tangga (KDRT), dan judi. Munculnya faktor judi dalam lima besar penyebab perceraian menunjukkan adanya pola yang unik. Hal ini mencerminkan perbedaan budaya antarwilayah, di mana konsumsi alkohol menjadi pemicu perceraian. Faktor meninggalkan pasangan juga berkaitan erat dengan fenomena migrasi internal. Fenomena ini juga didukung oleh budaya merantau yang kuat di beberapa daerah di Indonesia, di mana merantau dianggap sebagai upaya meningkatkan taraf hidup keluarga, namun di sisi lain dapat menyebabkan disintegrasi keluarga dan perubahan perilaku sosial [14].

Pada *cluster* 1, kasus perceraian yang tercatat jauh lebih besar dibandingkan *cluster* lain, dan faktor ekonomi menjadi penyebab kedua paling dominan. Hal ini menunjukkan bahwa di wilayah yang termasuk dalam *cluster* ini, tekanan ekonomi memiliki dampak signifikan terhadap stabilitas rumah tangga. Ketidakstabilan ekonomi, seperti pengangguran, inflasi, dan tingginya biaya hidup, sering kali memicu perselisihan dalam rumah tangga yang berujung pada perceraian [15].

Secara keseluruhan, perselisihan merupakan penyebab utama di semua *cluster*. Perselisihan merupakan keadaan dimana hal ini biasa disebabkan oleh perbedaan pendapat antara seorang istri dan suami, mulai dari perbedaan dalam pengurusan rumah tangga, perselisihan masalah nafkah dan lain sebagainya [16]. Sehingga, hal-hal seperti ini menyebabkan konflik dalam rumah tangga seringkali tidak terselesaikan dan berujung pada pertengkarannya yang berkepanjangan. Perbedaan faktor penyebab antar *cluster* mencerminkan variasi sosial, ekonomi, dan budaya antarwilayah. Di wilayah bertekanan ekonomi tinggi seperti Jawa, perceraian dipicu oleh masalah finansial, sementara di daerah lain lebih terkait judi dan konsumsi alkohol. Faktor meninggalkan salah satu pihak sering dikaitkan dengan budaya merantau dan urbanisasi, yang telah mengubah gaya hidup dan perilaku masyarakat. Secara keseluruhan, *clustering* dengan *DBSCAN* memberikan wawasan yang berbeda dari sisi jumlah *cluster*, karakteristik, identifikasi *outlier* dan interpretasi sosial dari data perceraian.

4. Kesimpulan

Penelitian ini telah berhasil mengelompokkan 34 provinsi di Indonesia berdasarkan pola penyebab perceraian menggunakan pemodelan *DBSCAN*, yang terpilih setelah sebelumnya

dilakukan perbandingan nilai Silhouette dengan model K-Means. Proses ini dilakukan berdasarkan data resmi yang didapat pada laman BPS Indonesia yang mencakup 13 faktor penyebab perceraian pada tahun 2024. Hasil dari pemodelan clustering memberikan wawasan yang bermanfaat dalam mengidentifikasi kelompok provinsi dengan karakteristik penyebab perceraian yang mirip. Metode DBSCAN menghasilkan dua cluster yang relatif seimbang, distribusi yang jelas dan interpretasi berdasarkan visualisasi serta analisis rata-rata tiap faktor. Sementara itu, metode K-Means kurang berhasil dalam mengelompokkan data secara optimal karena nilai Silhouette Score yang lebih rendah dari DBSCAN. Analisis terhadap cluster hasil DBSCAN menunjukkan bahwa, cluster 0 didominasi oleh faktor perselisihan dan pertengkarannya terus-menerus, meninggalkan salah satu pihak, ekonomi, kekerasan dalam rumah tangga (KDRT), dan judi. Di lain sisi, cluster 1 menunjukkan hampir sama, yaitu faktor perselisihan dan pertengkarannya terus-menerus, ekonomi, meninggalkan salah satu pihak, kekerasan dalam rumah tangga (KDRT), dan judi. Pemodelan ini mampu mendeteksi noise dengan lebih baik. Hasil penelitian ini memperlihatkan bahwa meskipun ada variasi antar provinsi, beberapa pola penyebab perceraian bersifat cukup konsisten dan menonjol secara nasional. Oleh karena itu, pendekatan kebijakan dan intervensi sosial perlu disesuaikan dengan karakteristik cluster, khususnya dalam mempertimbangkan pola sosial-budaya setiap cluster, agar program pencegahan dan penyuluhan dapat lebih tepat sasaran. Dengan demikian, clustering berbasis data statistik terbukti menjadi alat yang efektif untuk memahami fenomena sosial seperti perceraian secara lebih sistematis dan berbasis bukti.

Referensi

- [1] Badan Pusat Statistik Indonesia, "Jumlah Perceraian Menurut Provinsi dan Faktor Penyebab Perceraian (perkara), 2024," Badan Pusat Statistik Indonesia. Accessed: Jun. 01, 2025. [Online]. Available: <https://www.bps.go.id/id/statistics-table/3/YVdoU1IwVmITM2h4YzFoV1psWkViRXhqTlZwRFVUMDkjMw==/jumlah-perceraian-menurut-provinsi-dan-faktor-penyebab-perceraian--perkara---2024.html?year=2024>
- [2] Republik Indonesia, *UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 1 TAHUN 1974 TENTANG PERKAWINAN*. Indonesia: Badan Pemeriksa Keuangan, 1974. Accessed: Jun. 04, 2025. [Online]. Available: https://www.google.com/url?q=https://peraturan.bpk.go.id/Download/36382/UU%2520Nomor%252001%2520Tahun%25201974.pdf&sa=D&source=docs&ust=1749016417389455&usg=AOvVaw2S_bIumm8Tn2wekigRIBWm
- [3] N. Suryaningrum, "Determinan Perceraian di Jakarta Timur Tahun 2014 (Studi Data Pengadilan Agama dan Pengadilan Negeri)," *Forum Ilmu Sosial*, vol. 46, no. 2, pp. 128–141, 2019, doi: 10.15294/fis.v46i2.19627.
- [4] A. Tristanto, "PERCERAIAN DI MASA PANDEMI COVID-19 DALAM PERSPEKTIF ILMU SOSIAL," *Sosio Informa*, vol. 6, no. 3, 2020, doi: <https://doi.org/10.33007/inf.v6i3.2417>.
- [5] N. S. Manna, S. Doriza, and M. Oktaviani, "Cerai Gugat: Telaah Penyebab Perceraian Pada Keluarga di Indonesia," *Jurnal AL-AZHAR INDONESIA SERI HUMANIORA*, vol. 6, no. 1, 2021, doi: <https://doi.org/10.36722/sh.v6i1.443>.
- [6] S. Mutiah, Y. Hasnataeni, A. Fitrianto, E. Erfiani, and L. M. R. D. Jumansyah, "Perbandingan Metode Klustering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat," *Teorema: Teori dan Riset Matematika*, vol. 9, no. 2, p. 247, Sep. 2024, doi: 10.25157/teorema.v9i2.16290.
- [7] A. S. Devi, I. K. G. D. Putra, and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, p. 185, Dec. 2015, doi: 10.24843/lkjiti.2015.v06.i03.p05.
- [8] S. F. S. Reza and W. Cholil, "Implementasi Algoritma Random Forest Terhadap Prediksi Good Loan/Bad Loan Kredit Nasabah Bank Di Jakarta," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 4, no. 2, 2023, Accessed: Jun. 06, 2025. [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2578/1947>

-
- [9] E. N. R. Khakim, A. Hermawan, and D. Avianto, "IMPLEMENTASI CORRELATION MATRIX PADA KLASIFIKASI DATASET WINE," *JIKO (Jurnal Informatika dan Komputer)*, vol. 7, no. 1, p. 158, Feb. 2023, doi: 10.26798/jiko.v7i1.771.
 - [10] F. M. Pranata, S. H. Wijoyo, and N. Y. Setiawan, "Analisis Performa Algoritma K-Means dan DBSCAN Dalam Segmentasi Pelanggan Dengan Pendekatan Model RFM," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, pp. 2548–964, 2017, Accessed: Jun. 06, 2025. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/13962/6211>
 - [11] M. A. CHERID, "METODE DBSCAN CLUSTERING UNTUK ANALISIS POLA PENYEBARAN HUJAN DI SUMENEP," UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM, MALANG, 2023. Accessed: Jun. 06, 2025. [Online]. Available: <http://etheses.uin-malang.ac.id/59686/1/19610011.pdf>
 - [12] M. S. Badawi, F. Suria, E. Elfiani, W. Tsunami, and R. Saputra, "ANALISIS KLASTER PASIEN RUMAH SAKIT JIWA KOTA KENDARI TAHUN 2022-2023 IDENTIFIKASI POLA DAN KARAKTERISTIK MENGGUNAKAN METODE DBSCAN," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, pp. 3072–3079, Mar. 2025, doi: 10.36040/jati.v9i2.13247.
 - [13] A. Tate, "When To Choose Density-Based Methods (Compare, k-means, DBSCAN and Hierarchical Clustering)," Hex Technologies Inc. Accessed: Jun. 04, 2025. [Online]. Available: <https://hex.tech/blog/comparing-density-based-methods/#:~:text=DBSCAN%20inherently%20identifies%20and%20separates,susceptible%20to%20noise%20and%20outliers>.
 - [14] M. I. Sholik, F. Rosyid, K. Mufa'idah, T. Agustina, and U. R. Ashari, "MERANTAU SEBAGAI BUDAYA (EKSPLORASI SISTEM SOSIAL MASYARAKAT PULAU BAWEAN)," *Jurnal Cakrawala*, vol. 10, no. 2, 2016, Accessed: Jun. 01, 2025. [Online]. Available: <https://www.cakrawalajournal.org/index.php/cakrawala/article/view/39>
 - [15] M. Zakih, "Pengaruh Faktor Ekonomi dalam Kasus Perceraian Terhadap Putusan Pengadilan Agama di Jember (Studi Kasus Putusan PA Jember Nomor 4318/Pdt.G/2023/PA.Jr.)," *Jurnal Cahaya Mandalika*, vol. 5, no. 2, 2024, doi: <https://doi.org/10.36312/jcm.v5i2.3790>.
 - [16] W. M. Prananta and Ifrohati, "Pertengkar Terus Menerus Sebagai Alasan Paling Tinggi pada Perkara Cerai Gugat di Pengadilan Agama Kelas Ib Sungailiat," *Journal of Sharia and Legal Science*, vol. 2, no. 2, pp. 194–209, Aug. 2024, doi: 10.61994/jsls.v2i2.675.