CIPASUNG TECHNO PESANTREN
Scientific Journal
p-ISSN: 1978-0842 e-ISSN 2963-1114

Penerapan *Clustering* Pada Data Perceraian di Indonesia Menggunakan Algoritma BIRCH

Adittia Fathah[1,*]

[1]*Program Studi Informatika, Sekolah Tinggi Teknologi Cipasung*

[*]*Corresponding author's email*: adittiafathah@sttcipasung.ac.id

Diterima: DD-MM-YYYY Direvisi: DD-MM-YYYY Diterima setelah revisi: DD-MM-YYYY

Abstract. *Divorce is one of the social issues that has an impact on life, from adults to children. Divorce is influenced by many factors, ranging from economics, incompatibility, to the availability of technology. Therefore, appropriate efforts are needed to reduce the divorce rate. One approach that can be used is clustering using the BIRCH* (*Balanced Iterative Reducing and Clustering Using Hierarcics*) *algorithm based on divorce factors across all provinces in Indonesia. The study utilised 10 clusters, ranging from cluster 0 to 9. The research method employed was KDD. Data processing was conducted using Google Colab with the Python programming language. Based on the clustering results, different dominant factors were identified for each cluster. Cluster 2 has the most members, with 6 provinces, where the causes of divorce are abandonment by the spouse and gambling. This study produced a system for applying the BIRCH algorithm to divorce data, which can serve as a reference for the government to reduce divorce rates based on triggering factors in each region*.

Keywords: Clustering; BIRCH; Divorce; PCA

Abstrak. Perceraian adalah salah satu isu sosial yang memberi dampak dalam kehidupan, mulai dari orang dewasa hingga anak-anak. Perceraian dipengaruhi banyak faktor mulai dari ekonomi, ketidakcocokan hingga keberadaan teknologi. Karena itu, perlu adanya upaya yang sesuai agar angka perceraian dapat ditekan. Cara yang dapat digunakan adalah dengan melakukan *clustering* dengan algoritma BIRCH (*Balanced Iterative Reducing and Clustering Using Hierarcics*) berdasarkan faktor perceraian di seluruh provinsi di Indonesia. *Cluster* yang digunakan pada penelitian ini ada 10, mulai dari *cluster* 0 hingga 9. Metode Penelitian yang digunakan adalah KDD (*Knowledge Discovery in Database*). Pengolahan data dilakuan menggunakan Google Colab dengan bahasa pemroraman python. Berdasarkan hasil *cluster*, diperoleh faktor dominan yang berbeda-beda untuk setiap *cluster*nya. *Cluster* 2 memiliki anggota terbanyak yaitu 6 provinsi dengan penyebab perceraian karena ditinggalkan pasangan dan judi. Penelitian ini menghasilkan sistem penerapan algoritma BIRCH pada data perceraian, sehinga bisa menjadi salah satu acuan bagi pihak pemerintah untuk menekan angka perceraian berdasarkan faktor pemicu di setiap wilayah.

Kata kunci: Clustering; BIRCH, Perceraian; PCA

I. PENDAHULUAN

Perceraian adalah salah satu isu soasial yang memberi dampak dalam kehidupan, mulai dari orang dewasa hingga anak-anak [1]. Di Indonesia angka perceraian di tahun 2024 mencapai 399.921 kasus yang tersebar di berbagai provinsi di indonesia [12].Perceraian dapat dipicu oleh berbagai faktor seperti masalah ekonomi, kekerasan dalam rumah tangga, ketidakcocokan pasangan hingga perselisihan yang terus terjadi berulang kali [2]. Fenomena ini perlu menjadi perhatian khusus bagi permerintah atau lembaga sosial untuk mencari solusi yang bisa menekan tingginya angka perceraian. [11]. Selain masalah-masalah tadi, perceraian juga dapat dipengaruhi oleh perkembangan teknologi. Salah satunya adalah judi *online*. Sudah menjadi rahasia umum banyak pernikahan yang kandas akibat pengaruh judi *online* baik secara langsung maupun tidak [10]. Dalam kondisi tersebut, korban yang paling terdampak adalah anak-anak karena secara psikologis masih sangat rentan.

Menurut [3] terdapat beberapa efek kesehatan akibat perceraian diantaranya menimbulkan stress hingga depresi. Stress yang dialami bahkan dapat meningngkatkan resiko pada penyakit jantung dan *stroke*.

Perceraian tidak hanya meninggalkan luka psikologis bagi pasangan yang berpisah, tapi juga dapat menyebabkan trauma bagi anak-anak pasangan tersebut. Jika kondisi tersebut terus dibiarkan, maka akan berakibat negatif pada upaya mempersiapkan generasi penerus bangsa yang siap menghadapi tantangan hidup di masa depan.

Memahami faktor-faktor yang mempengaruhi perceraian tentu tidak mudah. Data perceraian seringkali sangat kompleks dengan banyaknya variabel yang berkaitan [11]. Diperlukan pendekatan khusus untuk memperoleh informasi dari data tersebut. Pendekatan yang bisa digunakan adalah *clustering*. Misalnya pada penelitian yang dilakukan oleh [1] dan [10] menggunakan *clustering* dengan algoritma K-Means untuk mengelompokkan provinsi di indonesia berdasarkan penyebab perceraian kedalam 3 *cluster*. Selain itu, [2] dan [11] juga menggunakan Algotrima K-Means untuk mengelompokan kasus perceraian dengan cakupan tingkat propinsi dengan jumlah *cluster* masing-masing 3 dan 4 *cluster*.

Berdasarkan pemaparan tersebut, penelitian dilakukan dengan tujuan untuk mengelompokkan data perceraian setiap provinsi di indonesia dengan pendekatan *clustering*. Namun, penelitian ini akan menggunakan algoritma yang berbeda yaitu BIRCH (*Balanced Iterative Reducing and Clustering Using Hierarcics*). Sedangkan untuk jumlah *cluster* akan dibuat berbeda juga yaitu 10 *cluster* untuk mendapatkan kemiripan faktor perceraian yang lebih spesifik untuk setiap *cluste*nya. Diharapkan adanya penelitian ini dapat memberikan gambaran yang lebih jelas mengenai kemiripan faktor perceraian di setiap provinsi sehingga angka perceraian dapat ditekan.

II. TINJAUAN PUSTAKA

2.1 Data Mining

Data Mining adalah proses untuk menemukan korelasi dan pola baru yang bermanfaat melalui eksplorasi dalam *repository* data yang besar. Teknologi pengenalan pola biasanya menggunakan teknik statistik [4]. *Data Mining* merupakan salah satu langkah dalam *Knowledge Discovery in Databese* (KDD). KDD sendiri terdiri atas pembersihan data (*cleaning*), integrasi data (*data integration*), pemilihan data (*data selection*), transformasi data (*data transformation*), *data mining*, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan (*knowledge*) [5].

2.2 *Clustering*

Clustering dikenal sebagai *segmentation*. Metode ini mengidentifikasi kelompok dalam sebuah kasus yang didasarkan pada kelompok atribut yang memiliki kemiripan. *Clustering* bekerja dengan memisahkan kelompok data berdasarkan ciri masing-masing. Objek *clusterig* dapat berupa orang atau peristiwa yang didistribusikan kedalam kelompok sehingga terdapat beberapa tingkatan yang saling berhubungan antar *cluster*. Kuat dan lemahnya antar anggota dari *cluster* yang berbeda terlihat pada anggota *cluster* yang sama [5].

2.3 BRICH

BIRCH (*Balanced Iterative Reducing and Clustering Using Hierarcics*) adalah algoritma *clustering* yang dirancang untuk menangani dataset berskala besar dan efisien dalam hal penggunaan memori dan waktu komputasi [6]. Algoritma ini memiliki keunggulan dalam menangani data yang besar dan akurat karena dapat mengatasi sebuah *outlier* atau pencilan. Algoritma BIRCH mempunyai dua cara kerja yatu *birc scan* atau membaca, dimana data yang dibangun akan masuk kedalam sebuah inisial memory CF *Tree* yang akan dipandang sebagai kompres multilevel dari data. Sedangkan cara kerja kedua adalah dengan menggunakan sebuah penyeleksian terhadap algoritma yang akan dipakai dalam bentuk *cluster leaf node* dari CF *tree* [7]. Proses ini memungkinkan BIRCH untuk melakukan pengelompokan bertahap dan efektif pada data besar tanpa perlu memuat seluruh data ke dalam memori. Algoritma ini cocok untuk data berdimensi tinggi dan berkinerja baik ketika jumlah kluster tidak terlalu besar dan distribusi data relatif seimbang [6].

2.4 PCA

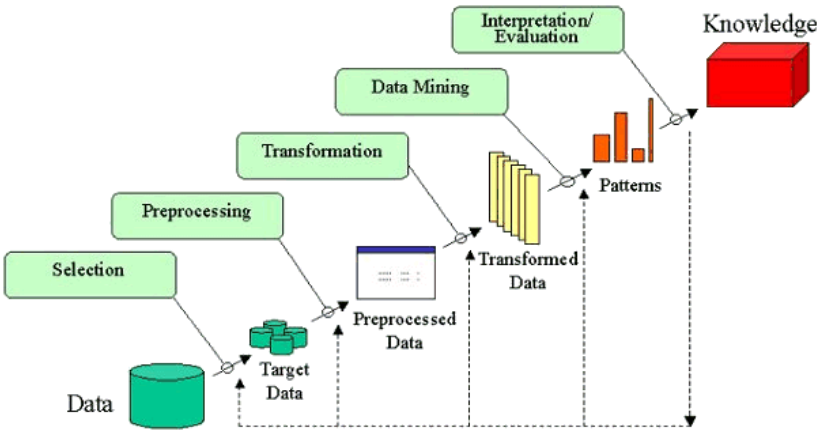
PCA (*Principal Component Analysis*) adalah metode yang digunakan untuk mengurangi dimensi pada dataset dengan merangkum variabel berdimensi tinggi menjadi kebih sedikit tanpa kehilangan informasi utama. Selain itu, PCA dapat membantu menguji variabel apakah dalam dataset memiliki kaitan atau independen seingga bisa dianalisis lebih lanjut. [8]. PCA merupakan bagian dari keilmuan statistika.

Penggunaan PCA tidak hanya pada keperluan statistika saja, tapi juga digunakan pada pengolahan citra yang dalam hal ini bekerja pada bagian *feature extraction* [9].

PCA umumnya digunakan ketika terjadi multikolinearitas pada data, yaitu ketika dua atau lebih variabel independen saling berkorelasi secara signifikan. Pengecekan multikolinearitas dapat dilakukan dengan beberapa cara, salah satunya adalah dengan visualisasi heatmap correlation berdasarkan koefisien Pearson. Heatmap adalah teknik visualisasi yang menyajikan kekuatan hubungan antar variabel dalam bentuk warna [8].

III. METODE PENELITIAN

Teknik yang digunakan dalam penelitian ini adalah KDD (K*nowledge Discovery in Database*). Tujuannya adalah untuk mengidentifikasi pola atau informasi penting dari data yang mentah menjadi pengetahuan yang berguna bagi pengambilan keputusan. Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap *pre-processing*, transformasi, *data mining* dan tahap evaluasi [4].



Gambar 1. Proses KDD.

IV. HASIL DAN PEMBAHASAN

4.1 Data yang Digunakan

Dataset yang menjadi dasar penelitiajn ini diperoleh dari *website* Badan Pusat Statistik (BPS). Data yang diambil adalah data perceraian di indonesia pada tahun 2024. Semua atributnya bertipe data *object*. Dataset diperoleh dalam format CSV dan terdiri atas 14 atribut. Semua atribut yang ada dapat diihat pada Gambar 2.

```
 0   Provinsi
 1   Fakor Perceraian - Zina
 2   Fakor Perceraian - Mabuk
 3   Fakor Perceraian - Madat
 4   Fakor Perceraian - Judi
 5   Fakor Perceraian - Meninggalkan Salah satu Pihak
 6   Fakor Perceraian - Dihukum Penjara
 7   Fakor Perceraian - Poligami
 8   Fakor Perceraian - Kekerasan Dalam Rumah Tangga
 9   Fakor Perceraian - Cacat Badan
 10  Fakor Perceraian - Perselisihan dan Pertengkaran Terus Menerus
 11  Fakor Perceraian - Kawin Paksa
 12  Fakor Perceraian - Murtad
 13  Fakor Perceraian - Ekonomi
 14  Fakor Perceraian - Jumlah
dtypes: object(15)
memory usage: 4.6+ KB
None
```

Gambar 2. Atribut pada Dataset.

4.2 *Data Selection*

Pada tahap ini, semua atribut digunakan, namun untuk keperluan *clustering*. Atribut provinsi dipisahkan dulu, karena datanya berisi huruf. Caranya adalah dengan menyalin kolom provinsi untuk disimpan, lalu atribut provinsi yang masih menyatu dihapus. Kode programnya dapat dilihat pada Gambar 3.

```
# Simpan Provinsi asli & hapus dari dataset sementara
provinsi_asli = data_cerai['Provinsi'].copy()
hapus_provinsi = data_cerai.drop(columns=['Provinsi'])
```

Gambar 3. Pemisahan Atribut Provinsi.

4.3 *Preprocessing* Data

Pada tahap *prepreocessing,* dilakukan pengecekan *missing value* untuk mengetahui apakah ada data yang kosong. Kode python yang digunakan dapat dilihat pada Gambar 4.

```
# Cek Missing Value
display(data_cerai.isnull().sum())
```

Gambar 4. Pengecekan *Missing Value*.

Sedangkan hasil pengecekannya dapat dilihat pada Gambar 5.

Provinsi	0
Fakor Perceraian - Zina	2
Fakor Perceraian - Mabuk	1
Fakor Perceraian - Madat	10
Fakor Perceraian - Judi	0
Fakor Perceraian - Meninggalkan Salah satu Pihak	0
Fakor Perceraian - Dihukum Penjara	0
Fakor Perceraian - Poligami	3
Fakor Perceraian - Kekerasan Dalam Rumah Tangga	0
Fakor Perceraian - Cacat Badan	5
Fakor Perceraian - Perselisihan dan Pertengkaran Terus Menerus	0
Fakor Perceraian - Kawin Paksa	13
Fakor Perceraian - Murtad	1
Fakor Perceraian - Ekonomi	0
Fakor Perceraian - Jumlah	0

Gambar 5. Hasil Pengecekan Missing Value.

Berdasarkan hasil pada Gambar 5, ada beberapa data yang memiliki *missing value*, yaitu faktor zina, mabuk, madat, cacat badan, kawin paksa dan murtad. Namun pada penelitian ini tidak dilakukan perbaikan data manual, karena di *website* BPS sudah disampaikan bahwa ada beberapa data yang tidak tercatat. Penyebabnya karena memang tidak ada kejadiannya untuk beberapa provinsi tertentu.

4.4 *Data Transformation*

Pada tahap ini, *encoding* dilakukan pada dataset. Tujuannya agar atribut yang memiliki tipe data *object* diubah menjadi tipe data angka. Kode yang digunakan dapat dilihat pada Gambar 6.

```
# Label Encoding kolom object
for col in data_cerai.columns:
    if data_cerai[col].dtype == 'object':
        encoder = LabelEncoder()
        data_cerai[col] = encoder.fit_transform(data_cerai[col].astype(str))
```

Gambar 6. Proses *Encoding*.

Sedangkan untuk format file, perubahan tidak dilakukan karena bawaannya sudah berfoemat CSV (*Comma Separated Value*).

4.5 *Data Mining*

Pada tahap *data mining*, teknik *clustering* yang digunakan adalah BIRCH dengan menggunakan library Birch. Sesuai dengan rencana diawal, jumlah *cluster* diatur ke angka 10 dan *threshold* 0.5. Kode programnya dapat dilihat pada Gambar 7.

```
# Jalankan BIRCH dengan threshold terbaik
birch_final = Birch(n_clusters=10, threshold=0.5)
labels_final = birch_final.fit_predict(data_cerai)
```

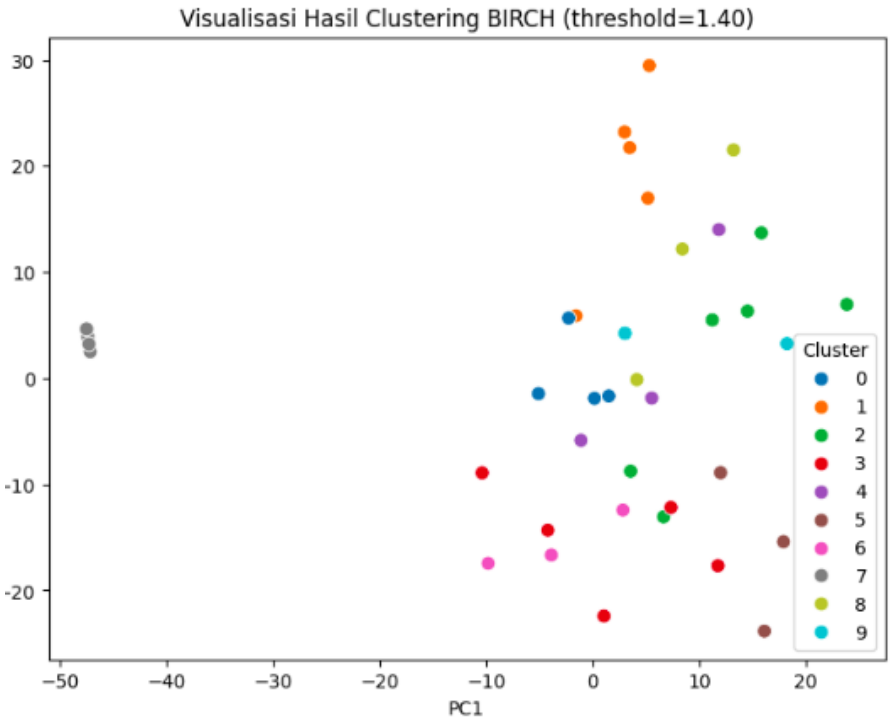
Gambar 7. Proses *Clustering* (BRICH).

Pengaturan n=10 dilakukan agar setelah proses pembentukan CF tree, BIRCH akan melakukan pengelompokan akhir hingga 10 *cluster*. Sedangkan *threshold* diatur ke 0.5 digunakan untuk menentukan jarak maksimum yang dianggap satu *cluster*, jika jaraknya lebih dari itu, *cluster* baru akan terbentuk. Hasil *cluster* akan disimapan dalam variabel *labels_final*. Setelah proses *cluster* selesai, atribut provinsi yang dipisah pada tahap awal, digabungkan lagi menjadi satu. Kode programnya dapat dilihat pada Gambar 8.

```
# Buat DataFrame hasil clustering
df_hasil = data_cerai.copy()
df_hasil['Provinsi'] = provinsi_asli
df_hasil['Cluster'] = labels_final
```

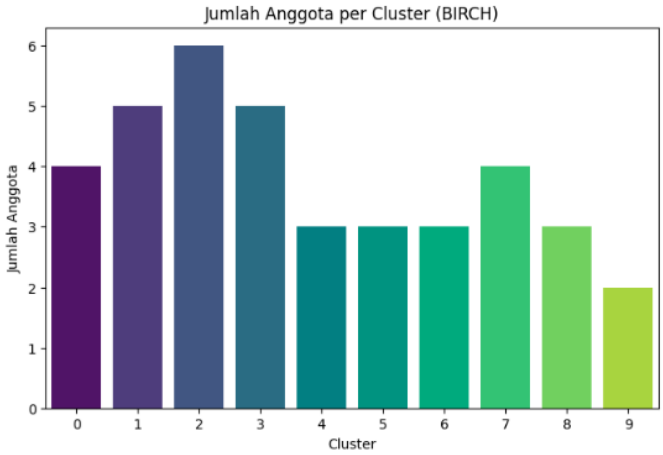
Gambar 8. Penggabungan Kembali Fitur Provinsi.

Untuk melihat lebih jelas sebaran *cluster* dalam format dua dimensi dilakukan visualisasi dengan PCA. Hasil sebarannya dapat dilihat pada Gambar 9.



Gambar 9. Sebaran Data Hasil *Clustering*.

Sedangkan untuk jumlah anggota setiap clusternya dapat dilihat pada Gambar 10.



Gambar 10. Diagram Jumlah Anggota Setiap *Cluster*.

Gambar 10 menunjukkan *cluster* 2 memiliki jumlah provinsi terbanyak yaitu 6, sedangkan *cluster* 9 memiliki jumlah provinsi paling sedikit yaitu 2. Hasil lengkap untuk setiap cluster dapat dilihat pada Tabel 1.

Tabel 1. Data Provinsi Hasil Clustering.

Cluster	**Anggota Provinsi**
0	a. Kepulauan Bangka Belitung b. Nusa Tenggara Barat, c. Nusa Tenggara Timur d. Kalimatan Utara
1	a. Sumatera Utara b. Lampung c. Sulawesi Utara d. Sulawesi Tengah e. Sulawesi Tenggara
2	a. Jawa Barat b. Jawa Tengah c. Jawa Timur d. Kalimantan Barat e. Kalimantan Timur f. Papua Barat
3	a. Jambi b. Bengkulu c. Kepulauan Riau d. Kalimanatan Tengah e. Gorontalo
4	a. Sumater Selatan b. Maluku Utara c. Papua
5	a. DI Yogyakarta b. Bali c. Kalimantan Selatan
6	a. Aceh b. DKI Jakarta c. Banten
7	a. Papua Barat Daya b. Papua Selatan c. Papua Tengah d. Papua Pegunungan
8	a. Sumatera Barat b. Riau c. Maluku

Cluster	Anggota Provinsi
9	a. Sulawesi Selatan b. Sulawesi Barat

4.6 Interpretasi Hasil

Interprestasi hasil dipaparkan berdasarkan hasil *clustering* yang dilakukan dengan algoritma BIRCH. Penelitian ini menekankan bahwa pendekatan berbasis data mampu mengelompokan setiap provinsi di indonesia berdasarkan kemiripan faktor cerai yang paling dominan, sehingga penanganannya dapat disesuaikan. Berikut adalah penjelasan lengkap dari masing-masing *cluster*.

1. *Cluster* 0 : Provinsi di *cluster* ini cenderung memiliki kasus perceraian yang tinggi akibat penyalahgunaan zat (madat/mabuk), sementara alasan seperti meninggalkan pasangan atau kawin paksa jarang sekali terjadi.
2. Cluster 1 : Provinsi di *cluster* ini, faktor ekonomi dan mabuk menjadi penyumbang terbesar perceraian, sedangkan konflik berkepanjangan jarang jadi alasan utama.
3. *Cluster* 2 : Provinsi di *cluster* ini, ditinggalkan pasangan dan judi menjadi penyebab utama perceraian, sementara kawin paksa dan alasan keagamaan jarang.
4. *Cluster* 3 : Provinsi di *cluster* ini, KDRT dan poligami menonjol sebagai penyebab perceraian, sementara faktor ekonomi dan zina relatif rendah.
5. *Cluster* 4 : Provinsi di *cluster* ini, perceraian banyak dipicu konfilik rumah tangga dan masalah ekonomi, sedangkan judi dan alasan agama jarang terjadi.
6. *Cluster* 5 : Provinsi di *cluster* ini, kekerasan dan kasus hukum berperan besar, sementara alasan kesehatan fisik jarang terjadi.
7. Cluster 6 : Provinsi di *cluster* ini, perceraian banyak dipicu oleh konflik berkepanjangan, zina dan pindah agama.
8. *Cluster* 7 : Provinsi di *cluster* ini, data perceraiannya 0. Hal ini terjadi karena data belum tercatat.
9. *Cluster* 8 : Provinsi di *cluste*r ini, perceraian didominasi karena pasangan meningggalkan rumah dan konflik, sedangkan poligami dan KDRT rendah.
10. *Cluster* 9 : Provinsi di *cluster* ini, perceraian didominasi karena konflik berkepanjangan, ekonomi dan poligami.

V. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan, hasil yang di dapatkan adalah sebagai berikut :

1. Penelitian ini menghasilkan sistem penerapan algoritma BIRCH pada data perceraian, sehinga bisa menjadi salah satu acuan bagi pihak pemerintah dalam upaya menekan angka perceraian berdasarkan faktor pemicu di setiap wilayah.
2. Algoritma BIRCH berhasil mengelompokan data perceraian dengan faktor dominan yang berebeda-beda di setiap clusternya.
3. Cluster 2 memiliki anggota jumlah provinsi yang paling banyak, artinya penyebab perceraian di Indonesia paling banyak karena ditinggalkan pasangan dan judi.

DAFTAR PUSTAKA

[1] D. Andriani and S. I. Azmi, “K-Means Clustering Tingkat Perceraian dan Faktor Perceraian di Indonesia Tahun 2023,” vol. 1, no. 1, 2025.
[2] A. O. Sari and M. Iqbal, “Analisis Data Mining Terhadap Data Faktor Perceraian Di Sumatera Utara Dengan Metode K-Means Clusstering,” vol. 4, 2025.
[3] M. Mahanani, “KLASTERISASI DATA PERCERAIAN DI WILAYAH KABUPATEN LAMONGAN DENGAN MENGGUNAKAN METODE K-MEANS,” 2024.
[4] D. F. Angraeni, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, “PEMANFAATAN ALGORITMA K-MEANS DALAM ANALISIS DATA PENJUALAN TOKO BUYUNG UPIK JS DI LAZADA,” JITET, vol. 13, no. 2, Apr. 2025, doi: 10.23960/jitet.v13i2.6438.
[5] F. Febriansyah, “PENERAPAN ALGORITMA K-MEANS CLUSTERING DATA GIZI BALITA PADA UPTD PUSKESMAS BUMI AGUNG,” JITET, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4923.

[6] B. Triwijaya and S. Wibowo, “Performance Comparison of K-Means Algorithm and BIRCH Algorithm in Clustering Earthquake Data in Indonesia with Web-Based Map Visualization,” vol. 8, no. 1, 2025.
[7] J. Laurenso, D. Jiustian, F. Fernando, V. Suhandi, and T. H. Rochadiani, “Implementation of K-Means, Hierarchical, and BIRCH Clustering Algorithms to Determine Marketing Targets for Vape Sales in Indonesia,” JAIC, vol. 8, no. 1, pp. 62–70, July 2024, doi: 10.30871/jaic.v8i1.4871.
[8] C. A. Nugraha, M. A. Kesuma, O. I. Cahyani, M. Wati, and H. Haviluddin, “Pengelompokan Harga Cabai Rawit Berdasarkan Provinsi Menggunakan Principal Component Analysis dan K-Means,” JUKI : Jurnal Komputer dan Informatika, vol. 7, no. 1, pp. 80–88, Jan. 2025.
[9] M. I. Al-Arrafi and A. Ramadhanu, “IMPLEMENTASI METODE ALGORITMA PRINCIPAL COMPONENT ANALYSIS (PCA) DAN ALGORITMA K-NEAREST NEIGHBOR (KNN) DALAM KLASIFIKASI BUAH JAMBU MADU JAMBU MERAH DAN MANGGIS,” 2025.
[10] A. Asistyasari, Y. Nuryaman, A. Yudha, and B. Sudarsono, “Analasis Penyebaran Kasus Perceraian Di Provinsi-Provinsi Indonesia Menggunakan Algoritma K-Mean,” INSIT, vol. 3, no. 01, pp. 40–45, Feb. 2025, doi: 10.34005/insit.v3i01.4507.
[11] S. A. Wardani, N. B. Al Varuq, and H. T. Santoso, “Implementasi Data Minning Clustering Dalam Mengelompokan Kasus Perceraian Yang Terjadi Di Provinsi Jawa Timur Menggunakan Algoritma K-Means,” jami, vol. 6, no. 1, pp. 68–83, June 2025, doi: 10.46510/jami.v6i1.324.
[12] B.-S. Indonesia, “Jumlah Perceraian Menurut Provinsi dan Faktor Penyebab Perceraian (perkara), 2024 - Statistical Data.” Accessed: Aug. 13, 2025. [Online]. Available: https://www.bps.go.id/en/statistics-table/3/YVdoU1IwVmlTM2h4YzFoV1psWkViRXhqTlZwRFVUMDkjMw==/jumlah-perceraian-menurut-provinsi-dan-faktor.html?year=2024