



Analisis Penyebaran Kasus Perceraian Di Provinsi-Provinsi Indonesia Menggunakan Algoritma K-Mean

Ayuni Asistiyasari¹, Yosep Nuryaman², Afri Yudha³, Bibit Sudarsono³

^{1,2,4}Program Studi Teknologi Informasi Universitas Bina Sarana Informatika

³Program Studi Teknologi Informasi Universitas Darma Persada

Ayuni.Yas@bsi.ac.id ,

Abstract

Abstract: *Technological developments do not always have a positive impact on various groups of society, including children, adults and the elderly. One of the things that has a negative impact is the increase in divorce due to easy internet access for non-positive activities such as online gambling and cheating. According to data received from the Central Statistics Agency, there have been 408,347 divorce cases spread across 39 provinces. This research was conducted to determine the distribution of divorce cases using the k-mean algorithm. From the research that has been carried out, divorce cases can be divided into 4 clusters with each cluster numbering 20 provinces, 3 provinces, 5 provinces and 6 provinces. The results of evaluating centroid proximity are -17122790.208.*

Keywords: *Data Mining, K-Mean, Clustering, Divorce*

Abstrak

Perkembangan teknologi tidak selalu memberikan dampak positif bagi berbagai kalangan masyarakat, baik anak-anak, dewasa maupun orang tua. Salah satu hal yang menjadi dampak negative yaitu meningkatnya perceraian karena akses internet yang mudah untuk kegiatan tidak positif seperti judi online, hingga berselingkuh. Menurut data yang diterima dari Badan Pusat Statistika telah terjadi 408.347 kasus perceraian yang tersebar di 39 provinsi. Penelitian ini dilakukan untuk mengetahui penyebaran kasus perceraian tersebut menggunakan algoritma k-mean. Dari penelitian yang sudah dilakukan, kasus perceraian tersebut dapat dibagi menjadi 4 kluster dengan masing-masing kluster berjumlah 20 provinsi, 3 provinsi, 5 provinsi dan 6 provinis. Adapun hasil evaluasi kedekatan centroid yaitu sebesar -17122790.208.

Kata kunci: Data Mining, K-Mean, Klusterisasi, Perceraian

I. PENDAHULUAN

Perkembangan teknologi saat ini tidak melulu memberikan dampak positif bagi warga negara Indonesia, tetapi juga dapat memberikan dampak negative. Dampak negative yang yang ditimbulkan ada

dari hal kecil hingga yang besar yang mengakibatkan seseorang tersandung tindak criminal seperti judi onlie.

Selain itu sudah menjadi informasi umum banyak perkawinan yang kandas yang disebabkan oleh pengaruh judi online baik secara langsung maupun tidak

langsung. Efek perceraian dapat memberikan dampak yang luas bagi setiap keluarga, dari istri, suami, anak maupun anggota keluarga lainnya.

Perceraian yang terus meningkat setiap tahunnya menjadi ancaman yang serius bagi perkembangan Masyarakat Indonesia, karena dari keluargalah Pendidikan dalam bernegara bisa dikembangkan.

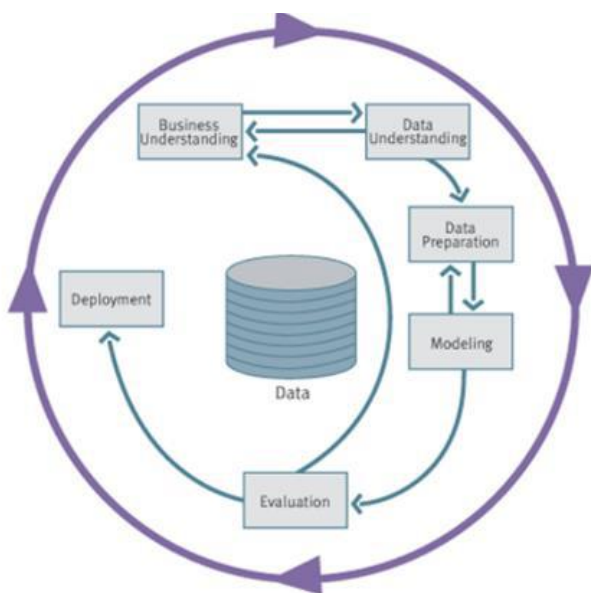
Algoritma K-Mean merupakan salah satu algoritma clustering non hierarki yang mempartisi data yang ada ke dalam betuk satu atau lebih atau kelompok. Sehingga kelompok yang memiliki karakteristik sama dikelompokkan dalam 1 kelompok dan data yang memiliki karakteristik yang berbeda akan dikelompokkan kedalam kelompok lainnya[1].

Algoritma K-Mean mampu mengelompokkan data obat-obatan di Rumah Sakit Umum Daerah Pekanbaru yang dapat digunakan sebagai referensi dalam pengambilan keputusan dalam perencanaan dan pengendalian pasokan medis di rumah sakit tersebut[2].

Oleh sebab itu kami mencoba melakukan penelitian tentang jumlah perceraian yang terjadi di provinsi di Indonesia dengan cara mengelompokkan provinsi-provinsi tersebut dalam beberapa cluster menggunakan algoritma k-mean[3].

II. METODE PENELITIAN

Me Penelitian ini menggunakan model CRISP-DM (Cross-Industry Standard Proses for Data Mining), yang terdiri dari 6 tahap yaitu



Gambar 1. CRISP-DM

1. Business/Research Understanding Phase
Perceraian yang semakin meningkat tentunya akan memberikan dampak negative. Dengan

membagi-bagi kota-kota di Indonesia kedalam kluster-kluster menggunakan algoritma k-mean diharapkan mampu memberikan gambaran bagaimana kota-kota tersebut dikelompokkan.

2. Data Understanding Phase
Data yang diambil adalah data jumlah perceraian masing-masing kota di Indonesia tahun 2023. Data tersebut diambil dari website Badan Pusat Statistik [4]. Terdapat 38 kota dengan 408.347 kasus perceraian yang terdata.

TABEL I
JUMLAH PERCERAIAN PER PROVINSI DI INDONESIA

Provinsi	Jumlah Kasus
Aceh	6.086
Sumatera Utara	15.660
Sumatera Barat	8.036
Riau	8.617
Jambi	4.573
Sumatera Selatan	10.374
Bengkulu	3.641
Lampung	14.038
Kep Babel	2.239
Kepulauan Riau	3.301
DKI Jakarta	14.381
Jawa Barat	91.146
Jawa Tengah	68.133
DI Yogyakarta	5.187
Jawa Timur	79.248
Banten	14.133
Bali	1.065
NTT	6.766
NTB	471
Kalbar	4.979
Kalimantan Tengah	3.147
Kalimantan Selatan	6.537
Kalimantan Timur	6.910
Kalimantan Utara	977
Sulawesi Utara	1.818
Sulawesi Tengah	3.581
Sulawesi Selatan	12.806
Sulawesi Tenggara	3.673
Gorontalo	2.062
Sulawesi Barat	1.151
Maluku	793

Maluku Utara	1.168
Papua Barat	494
Papua Barat Daya	
Papua	1.156
Papua Selatan	
Papua Tengah	
Papua Pegunungan	
Indonesia	408.347

23	Kalimantan Timur	6.910
24	Kalimantan Utara	977
25	Sulawesi Utara	1.818
26	Sulawesi Tengah	3.581
27	Sulawesi Selatan	12.806
28	Sulawesi Tenggara	3.673
29	Gorontalo	2.062
30	Sulawesi Barat	1.151
31	Maluku	793
32	Maluku Utara	1.168
33	Papua Barat	494
34	Papua	1.156
	Total	408.347

3. Data Preparation Phase

Data yang sudah didapatkan kemudian dilakukan preparation agar siap diolah menggunakan model yang ingin digunakan. Dari 39 kota yang diperoleh setelah tahap preparation menjadi 34 kota.

TABEL II
DATA SIAP OLAH PERCERAIAN PER
PROVINSI DI INDONESIA

No	Provinsi	Jumlah Kasus
1	Aceh	6.086
2	Sumatera Utara	15.660
3	Sumatera Barat	8.036
4	Riau	8.617
5	Jambi	4.573
6	Sumatera Selatan	10.374
7	Bengkulu	3.641
8	Lampung	14.038
9	Kep Bangka Belitung	2.239
10	Kepulauan Riau	3.301
11	DKI Jakarta	14.381
12	Jawa Barat	91.146
13	Jawa Tengah	68.133
14	DI Yogyakarta	5.187
15	Jawa Timur	79.248
16	Banten	14.133
17	Bali	1.065
18	Nusa Tenggara Barat	6.766
19	Nusa Tenggara Timur	471
20	Kalimantan Barat	4.979
21	Kalimantan Tengah	3.147
22	Kalimantan Selatan	6.537

4. Modeling Phase

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran computer (machine learning) untuk menganalisis dan mengecekstarki pengetahuan (knowledge) secara otomatis

Menambahkan definisi lain dari data mining merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sah (Sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database [1].

KDD (Knowledge discovery in database)

Istilah data mining dan KDD seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar [6]. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu hapan dalam keseluruhan proses KDD adalah data mining.

Klustering

Analisis kluster atau clustering merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang kesamaan datanya dalam suatu kelompok lebih besar daripada kesamaan data tersebut dengan data dalam kelompok lain[5].

K-MEAN

K-means clustering termasuk dalam partitioning clustering yang disebut juga exclusive clustering yang memisahkan data ke k daerah bagian yang terpisah dan setiap data harus termasuk ke dalam cluster tertentu dan memungkinkan setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke

cluster yang lain. K-means merupakan algoritma yang sangat terkenal karena kemudahan dan kemampuannya untuk melakukan pengelompokan data besar dengan sangat cepat [3].

Model yang digunakan yaitu k-mean. Kami menggunakan rapid miner untuk mempermudah pengolahannya.

5. Evaluation Phase

Untuk mengukur seberapa baik hasil kluster yang dihasilkan, kami melakukan evaluasi menggunakan metode pengukuran kedekatan centroid

6. Deployment Phase

Pada fase ini dilakukan implementasi hasil penelitian yaitu dengan menampilkan penyebaran klasterisasi tersebut

III. HASIL DAN PEMBAHASAN

Setelah persiapan dilakukan maka dilakukan perhitungan menggunakan K-Mean. Seperti langkah-langkah yang dijelaskan sebelumnya, langkah pertama yang dilakukan adalah menentukan titik centroid awal.

Penentuan titik centroid awal dilakukan random. Jumlah centroid juga menyesuaikan dalam hal ini yaitu 4 centroid. Untuk titik centroid dalam kasus ini seperti yang ada dibawah ini.

TABEL III
TITIK CENTROID

No	Provinsi	Jumlah Kasus
1	Aceh	6.086
2	Sumatera Utara	15.660
3	Sumatera Barat	8.036
4	Riau	8.617

Setelah ditentukan titik centroidnya, maka langkah selanjutnya yaitu menghitung jarak-jarak data terhadap titik centropdi yang sudah ditentukan sesuai dengan rumus yang sudah ada, Setelah diketahui jarak-jakar masing-masing data terhadap centroid yang sudah ditentukan, maka ditentukan data tersebut masuk ke centroid mana berdasarkan nilai jarak tersebut, semakin dekat jarak ke centroid.

TABEL IV
LITERASI 1 JARAK KE TITIK CENTROID

Data	Jarak ke Centroid				Kategori
	C1	C2	C3	C4	
1	0	4541	1553	838	C1
2	63	0	1555	724	C2
3	2124	4491	0	783	C3
4	778	4817	1243	0	C3
5	1499	3258	2085	836	C3

Setelah dilakukan perhitungan kedekatan dan penentuan titik centroid, maka didapatkan hasil literasi 1 untuk kategori centroid masing-masing provinsi. Dari hasil tersebut kemudian dicari letak centroid baru.

TABEL V
CENTROID BARU

	Nilai
C1	366
C2	2800
C3	1278

Setelah diketahui titik centroid baru maka dilakukan perhitungan untuk mencari kedekatan masing-masing data terhadap centroid baru tersebut seperti yang dilakukan pada literasi 1

TABEL VI
LITERASI 2 JARAK KE TITIK CENTROID

Data	Jarak ke Centroid				Kategori
	C1	C2	C3	C4	
1	144	3064	1914	1272	C1
2	97	2968	1905	1173	C1
3	117	3016	1911	1222	C1
4	776	3331	1752	1651	C1
5	1484	1893	2376	724	C3
6	3376	806	2950	2293	C2
7	4420	1488	4167	3269	C2
8	4108	1221	3962	2955	C2
9	118	2982	1787	1204	C1
10	3172	2261	2340	2534	C2

Sama halnya dengan proses sebelumnya, dilakukan pencarian untuk menentukan titik centroid baru. Namun perlu dilakukan perhitungan rasio untuk dari hasil literasi

pertama dan literasi awal dengan literasi kedua. Apabila rasio literasi pertama lebih kecil dari pada rasio literasi kedua, maka hitungan literasi dihentikan. Namun apabila hasil rasio literasi kedua masih lebih tinggi dari pada literasi pertama hitungan dilanjutkan.

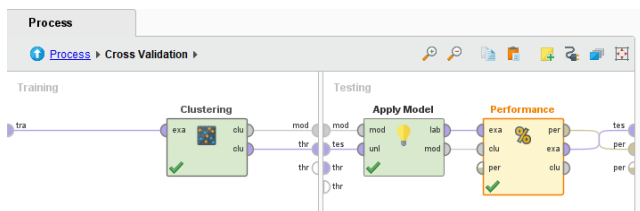
Dan pada kasus ini rasio literasi kedua masih lebih besar daripada hasil rasio literasi pertama, sehingga perhitungan dilanjutkan ke literasi ketiga. Seperti pada tahap sebelumnya, dilakukan perhitungan untuk mencari titik centroid yang baru.

TABEL VII
TITIK CENTROID BARU

	W
C1	384286
C2	3400000
C3	1142500

Dan dari hasil titik centroid baru tersebut, dilakukan literasi ke 3 yaitu menghitung jarak data-data ke titik centroid baru tersebut serta menggolongkan data-data yang ada ke centroid terdekat. Hal itu terus dilakukan sampai literasi-literasi selanjutnya sampai didapatkan nilai rasio literasi terakhir tidak lebih besar dari literasi sebelumnya.

Selain itu dilakukan perhitungan algoritma k-mean menggunakan rapid miner untuk mempermudah mencari hasil terakhir dari literasi tersebut



Gambar 2. Pengolahan data training

Dari hasil perhitungan yang dilakukan menggunakan rapid miner dihasilkan 4 klusterisasi terbaik yaitu

- Cluster 0 terdiri dari 20 kota yaitu kota 1,5,6,11,12,13,15,16,17,19,31,22,24,25,27,28,30,32,33,34
- Cluster 1 terdiri dari 3 kota yaitu kota 18,20,26
- Cluster 2 terdiri dari 5 kota yaitu kota 2,3,4,23,31
- Cluster 3 terdiri dari 6 kota yaitu yaitu kota 7,8,9,10,14 dan 29

Provinsi	Cluster	Jumlah
----------	---------	--------

Sumatera Utara	cluster_0	15.660
NTT	cluster_0	471
Sulawesi Tenggara	cluster_0	3.673
DI Yogyakarta	cluster_0	5.187
Kalimantan Barat	cluster_0	4.979
Kalimantan Selatan	cluster_0	6.537
Nusa Tenggara Barat	cluster_0	6.766
Maluku	cluster_0	793
Maluku Utara	cluster_0	1.168
Jawa Barat	cluster_0	91.146
Lampung	cluster_0	14.038
Sulawesi Selatan	cluster_0	12.806
Sumatera Barat	cluster_0	8.036
Kep Babel	cluster_0	2.239
Sulawesi Barat	cluster_0	1.151
Riau	cluster_0	8.617
Bengkulu	cluster_0	3.641
Bali	cluster_0	1.065
Sulawesi Tengah	cluster_0	3.581
Papua	cluster_0	1.156
Jambi	cluster_1	4.573
Sulawesi Utara	cluster_1	1.818
Jawa Tengah	cluster_1	68.133
Kalimantan Tengah	cluster_2	3.147
Papua Barat	cluster_2	494
DKI Jakarta	cluster_2	14.381
Gorontalo	cluster_2	2.062
Jawa Timur	cluster_2	79.248
Aceh	cluster_3	6.086
Kepulauan Riau	cluster_3	3.301
Kalimantan Timur	cluster_3	6.910
Kalimantan Utara	cluster_3	977
Banten	cluster_3	14.133
Sumatera Selatan	cluster_3	10.374

IV. KESIMPULAN DAN SARAN

Dari perhitungan yang sudah dilakukan maka dapat disimpulkan bahwa hasil terbaik yaitu terbagi menjadi 4 kluster yaitu kluster 0,1,2 dan 3. Dimana kluster 0 terdiri dari 20 kota, kluster 1 terdiri dari 3 kota, kluster 2 terdiri dari 5 kota dan kluster 3 terdiri dari 6 kota dengan nilai kedekatan centroid sebesar -17122790.208.

Adapun saran untuk penelitian selanjutnya yaitu pengklasifikasian hal-hal yang menyebabkan perceraian.

DAFTAR PUSTAKA

- [1] B. Sudarsono, U. Faddillah, A. Asistiyasari, and Y. Nuryaman, "Klasifikasi Data Penjualan Menggunakan Algoritma K-Means Dan Analytic Hierarchy Process," 2022. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/larik>
- [2] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [3] D. Bima, A. Pramana, T. Amalina, and R. I. Adam, "Analisis K-Means Clustering Pada Pengiriman Produk Bearing," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 128–137, 2022, doi: 10.5281/zenodo.7048988.
- [4] Badan Pusat Statistik, "Jumlah Perceraian Menurut Provinsi dan Faktor 2023," <https://www.bps.go.id/id/statistics-table/3/YVdoU1IwVmlTM2h4YzFoV1psWkViRXhqTlZwRFVUMDkjMw==/jumlah-perceraian-menurut-provinsi-dan-faktor.html?year=2023>.