

Penerapan Algoritma K-Medoids Dalam Mengelompokkan Tingkat Kasus Kejahatan di Setiap Provinsi

Nur Arief^{1*}, Irfan Sudahri Damanik², Eka Irawan³

^{1,2,3}STIKOM Tunas Bangsa, Pematangsiantar, Sumatera Utara, Indonesia

Email: ¹arief21nur12@gmail.com, ²Irfansudahri@amiktunasbangsa.ac.id, ³ekairawanstikomtunasbangsa.ac.id

Email Penulis Korespondensi: arief21nur12@gmail.com

Abstrak—Penyalahgunaan narkoba merupakan masalah yang sangat serius dan perlu mendapatkan perhatian khusus dari semua pihak. Hal ini dibuktikan dengan semakin meningkatnya kasus narkoba dan kasus kematian akibat penyalahgunaan narkoba yang dilaporkan dari berbagai media. Dampak kecanduan narkoba dapat terlihat pada fisik, psikis maupun social seseorang. Kasus penyalahgunaan narkoba pada setiap provinsi di Indonesia memiliki tingkat kasus yang beragam. Penelitian ini bertujuan untuk mengetahui cluster kasus penyalahgunaan narkoba pada tingkat tinggi dan rendah. Metode yang dibutuhkan untuk pengelompokan data kasus narkoba yaitu menggunakan metode data mining dengan algoritma K-Medoids dan menggunakan sebuah sistem komputerisasi yaitu aplikasi rapidminer 5.3. Data yang digunakan bersumber dari Badan Narkotika Nasional Republik Indonesia dengan website:<https://bnn.go.id/data> tahun 2015 – 2017 yang terdiri dari 34 provinsi untuk dibagi dalam 2 cluster. Hasil dari perhitungan algoritma K-Medoids diperoleh cluster tinggi sebanyak 10 provinsi dan cluster rendah sebanyak 24 provinsi. Pengelompokan ini diharapkan menjadi masukan bagi pemerintah atau pihak terkait untuk lebih meningkatkan sosialisasi bahaya narkoba agar dapat meminimalisir angka kematian dan kriminalitas akibat narkoba. adalah penentuan probabilitas prior bagi tiap kategori berdasarkan sampel dokumen.

Kata Kunci: Penyalahgunaan Narkoba, K-Medoids, Data Mining

Abstract—Drug abuse is a very serious problem and needs special attention from all parties. This is evidenced by the increasing number of drug cases and death cases due to the purchase of drugs issued from various media. The impact of drug addiction can be seen in one's physical, psychological and social. Drug abuse cases in each province in Indonesia have varying degrees of cases. This study aims to determine the cluster of drug abuse at high and low levels. The method needed for grouping drug case data is using data mining methods with the K-Medoids algorithm and using a computerized system that is rapidminer 5.3 application. The data used is sourced from the National Narcotics Agency of the Republic of Indonesia with the website: <https://bnn.go.id/2015-2017> data which consists of 34 provinces to be divided into 2 clusters. From the calculation of the K-Medoids algorithm, high clusters were 10 provinces and low clusters were 24 provinces. This grouping is expected to be included for the government or related parties to further increase the socialization of the dangers of drugs in order to minimize mortality and crime due to drugs.

Keywords: Drug Abuse, K-Medoids, Data Mining

1. PENDAHULUAN

Proses modernisasi dari evolusi budaya atau adat merupakan perubahan yang umum terjadi pada sebuah negara. Indonesia merupakan negara berkembang yang dalam proses modernisasinya secara tidak langsung membuat perubahan dalam berbagai hal, antara lain budaya, adat dan pola pikir masyarakat. Proses ini dapat berdampak negatif yang menarik perhatian masyarakat seperti terjadinya kriminalitas atau kejahatan yang melibatkan generasi muda sebagai pelaku, seperti kejahatan yang terjadi yaitu kejahatan pembunuhan berencana yang dilakukan oleh remaja, kerabat dekat dan teman dekat. Negara Indonesia tetap mengakui dan mempertahankan pidana mati untuk menjerat orang yang melakukan kejahatan. Hal ini sesuai dengan ketentuan pidana, yang memuat penjatihan sanksi tindak pidana pencurian dengan kekerasan cukup tinggi yang terjadi di Indonesia.

Berdasarkan uraian diatas banyak cabang kecerdasan buatan dalam ilmu komputer yang dapat menyelesaikan permasalahan tersebut secara kompleks diantaranya sistem pendukung keputusan, sistem pakar, data mining dan lain sebagainya. Beberapa penelitian tentang data mining, salah satunya yang dilakukan oleh (Pramessti, Furqon & Dewi 2017) dalam penelitian ini metode yang digunakan dalam penelitian ini adalah k-medoids clustering yang dapat diimplementasikan pada pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas (hotspot) menghasilkan jumlah cluster dan jumlah data mempengaruhi terhadap hasil kualitas dari cluster berdasarkan proses pengujian yang dengan jumlah 2 cluster dan jumlah data 7352. Penelitian selanjutnya dilakukan oleh [2] Dalam penelitian ini dilakukan pengujian model dengan menggunakan K-Medoids Chebyshev dalam pengelompokan data EDGI E-government Surve 2014 kedalam 4 status EDGI yang menghasilkan model mendapatkan nilai Bouldin Index dari setiap algoritma yang lebih optimal dalam penentuan pengelompokan EDI E-governmnet Surve 2014 ke dalam 4 status EDGI di dibandingkan menggunakan mahattan atau ecludian. Penelitian ini menjadi salah satu acuan penulis dalam melakukan penelitian sehingga penulis dapat memperkaya teori yang digunakan dalam mengkaji penelitian yang dilakukan.

Untuk mengetahui tingkat kasus kejahatan di setiap provinsi penulis menggunakan teknik data mining dalam proses pengolahan data dengan algoritma k-medoids. Data yang telah didapatkan selanjutnya dikelompokkan kedalam cluster dan diuji menggunakan RapidMiner 5.3 pada persentase desa/kelurahan yang pernah ada kejadian kejahatan pencurian dengan kekerasan selama setahun terakhir menurut provinsi, tahun 2008, 2011, dan 2014. Dengan diterapkannya algoritma k-medoids dalam mengelompokkan tingkat kasus kejahatan di setiap provinsi, diharapkan dengan adanya penelitian ini dapat digunakan sebagai salah satu dasar pengambilan keputusan kepada pihak pemerintah

dalam upaya penanggulangan atau mengatasi seseorang agar tidak melakukan kejahatan. Serta diperlukan pembinaan dalam hal pendidikan moral, pendidikan norma agama dan bermasyarakat menurut provinsi yang berdampak menurunkan angka tingkat kasus kejahatan di setiap provinsi. Sehingga dapat mengurangi perbuatan yang bertentangan dengan agama atau bertentangan dengan undang-undang.

2. METODOLOGI PENELITIAN

2.1. Data Mining

Data mining adalah proses analitik yang dirancang untuk memeriksa sejumlah data yang besar dalam mencari suatu pengetahuan tersembunyi yang berharga dan konsisten. *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan diantaranya deskripsi, estimasi, prediksi, klasifikasi, pengelompokan dan asosiasi [3]. *Data mining* digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases* (KDD) yaitu tahapan yang dilakukan dalam menggali pengetahuan dari sekumpulan data. Proses penggalian informasi tersembunyi dalam suatu basis data yang besar sering menggunakan istilah *datamining* dan *knowledge discovery in databases* (KDD) [3].

2.2. Clustering

Clustering adalah salah satu teknik *datamining* yang bertujuan untuk mengidentifikasi sekelompok objek yang mempunyai kemiripan karakteristik tertentu yang dapat dipisahkan dengan kelompok objek lainnya, sehingga objek yang berada dalam kelompok yang sama relatif lebih homogen dari pada objek yang berada pada kelompok yang berbeda [4]. Tujuan dari pengelompokan sekumpulan data objek kedalam beberapa kelompok yang mempunyai karakteristik tertentu dan dapat dibedakan satu sama lainnya adalah untuk analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian yang dilakukan. Model yang diambil diasumsikan bahwa data yang dapat digunakan adalah data yang berupa data interval, frekuensi dan biner.

2.3. Metode K-Medoids Clustering

Menurut [5] Metode *k-medoids* adalah metode pengelompokan yang berkaitan dengan metode *k-means* dan metode *medoidshift*. Metode *k-medoids* di usulkan pada tahun 1987. Metode *k-medoids* merupakan teknik partisi klasik *clustering* yang mengelompokkan data set dari n objek ke dalam kelompok k dikenal apriori. Metode ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Objek yang terpilih untuk mewakili sebuah *cluster* disebut dengan *medoid*. *Cluster* di bangun dengan menghitung kedekatan yang dimiliki antara *medoid* dengan objek *non-medoid*. Langkah-langkah metode *k-medoids*:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(x_i - x_j)'(x_i - x_j)} \quad (1)$$

dimana $i=1, \dots, n; j=1, \dots, n$ dan p adalah banyak variable, serta V adalah matrik varian kovarian.

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.
5. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoid*.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoid*, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing.

3. HASIL DAN PEMBAHASAN

Data pada penelitian ini diperoleh melalui media perantara atau secara tidak langsung berupa bukti yang telah ada, atau arsip baik yang dipublikasikan maupun yang tidak dipublikasikan secara umum atau biasa disebut dengan data sekunder. Data sampel yang digunakan berupa nama-nama 34 provinsi di Indonesia serta tingkat tingkat persentase kejahatan pencurian dengan kekerasan selama setahun terakhir menurut provinsi, tahun 2008, 2011, dan 2014.

Tabel 1. Data Persentase Tingkat Kejahatan Pencurian Dengan Kekerasan

No	Provinsi	Tahun		
		2008	2011	2014
1	Aceh	2,18	0,97	0,72
2	Sumatera Utara	3,23	1,6	2,31
3	Sumatera Barat	6,17	5,32	6,2
4	Riau	9,29	4,77	5,5
5	Jambi	5,07	4,15	4,9
6	Sumatera Selatan	9,06	6,94	13,28

No	Provinsi	Tahun		
		2008	2011	2014
7	Bengkulu	2,81	2,19	2,61
8	Lampung	8,55	7,55	9,69
9	Kep. Bangka Belitung	6,98	4,43	3,94
10	Kep. Riau	6,44	3,68	4,1
11	DKI Jakarta	10,49	13,11	27,34
12	Jawa Barat	5,06	4,86	5,77
13	Jawa Tengah	1,54	1,7	2,05
14	DI Yogyakarta	1,6	6,39	4,57
15	Jawa Timur	3,16	3,41	4,93
16	Banten	5,19	3,52	3,16
17	Bali	1,12	1,82	3,21
...
...
31	Maluku	0	0,88	1,84
32	Maluku Utara	0,48	0,19	0,59
33	Papua Barat	0,25	0,9	0,83
34	Papua	1,95	3,9	2,32

Langkah-langkah penyelesaian menggunakan algoritma *K-Medoids* antara lain sebagai berikut:

1. Inisialisasi pusat *cluster* sebanyak dari data sampel. Untuk pemilihan setiap *medoid* dipilih secara acak (random). Diasumsikan Jawa Timur dan Banten sebagai *medoid* awal.

Tabel 2. Medoid Awal

Nama	2008	2011	2014
Jawa Timur(C1)	3,16	3,41	4,93
Banten(C2)	5,19	3,52	3,16

2. Menghitung nilai jarak terdekat (*cost*) dengan persamaan *Euclidian Distance* yakni persamaan (1). Maka perhitungan untuk menghitung jarak setiap objek dengan *medoid* awal adalah sebagai berikut:

$$D_{aceh, C1} = \sqrt{(2,18 - 3,16)^2 + (0,97 - 3,41)^2 + (0,72 - 4,93)^2} = 4,963678$$

$$D_{aceh, C2} = \sqrt{(2,18 - 5,19)^2 + (0,97 - 3,52)^2 + (0,72 - 3,16)^2} = 4,638556$$

⋮

$$D_{Papua, C1} = \sqrt{(1,95 - 3,16)^2 + (3,9 - 3,41)^2 + (2,32 - 4,93)^2} = 2,91827$$

$$D_{Papua, C2} = \sqrt{(1,95 - 5,19)^2 + (3,9 - 3,52)^2 + (2,32 - 3,16)^2} = 3,36862$$

Hasil dari keseluruhan dapat dilihat pada tabel 4.2 sebagai berikut :

Tabel 3. Hasil Perhitungan Algoritma *K-Medoids* Iterasi Ke-1

No	Provinsi	Jarak ke Medoid	
		C1	C2
1	Aceh	4,963678	4,638556
2	Sumatera Utara	3,185184	2,872368
3	Sumatera Barat	3,784323	3,666333
4	Riau	6,304871	4,883452
5	Jambi	2,04856	1,854427
6	Sumatera Selatan	10,81635	11,36168
7	Bengkulu	2,644485	2,781331
8	Lampung	8,297548	8,376837
9	Kep. Bangka Belitung	4,075893	2,154205
10	Kep. Riau	3,394142	1,572164
11	DKI Jakarta	25,49563	26,54676
12	Jawa Barat	2,533397	2,936767
13	Jawa Tengah	3,720605	4,226937
...
...
32	Maluku Utara	6,032114	6,314895
33	Papua Barat	5,619448	6,057797
34	Papua	2,91827	3,36862
Jumlah		172,84	172,141
TotalCost		344,9809702	

Setelah didapatkan hasil jarak dari setiap objek (*cost*) pada iterasi ke-1 maka lanjut ke iterasi ke-2. Kandidat *medoid* baru (*non medoid*) pada iterasi ke-2 dapat dilihat pada tabel 4 sebagai berikut :

Tabel 4. Medoid Baru (Non-Medoid) Iterasi Ke-2

Nama	2008	2011	2014
Bengkulu(C1)	2,81	2,19	2,61
Sulawesi Utara(C2)	1,47	1,59	1,14

Hitung kembali jarak dari setiap objek pada iterasi ke-2 dengan medoid baru menggunakan persamaan (1).

$$D_{aceh, C1} = \sqrt{(2,18 - 2,81)^2 + (0,97 - 2,19)^2 + (0,72 - 2,61)^2} = 2,336108$$

$$D_{aceh, C2} = \sqrt{(2,18 - 1,47)^2 + (0,97 - 1,59)^2 + (0,72 - 1,14)^2} = 1,03194$$

⋮

$$D_{Papua, C1} = \sqrt{(1,95 - 2,81)^2 + (3,9 - 2,19)^2 + (2,32 - 2,61)^2} = 1,935924$$

$$D_{Papua, C2} = \sqrt{(1,95 - 1,47)^2 + (3,9 - 1,59)^2 + (2,32 - 1,14)^2} = 2,637973$$

Maka didapatkan hasil dari keseluruhan iterasi ke-2 dapat dilihat pada tabel 5 sebagai berikut :

Tabel 5. Hasil Perhitungan Algoritma K-Medoids Iterasi Ke-2

No	Provinsi	Jarak ke Medoid	
		C1	C2
1	Aceh	2,336108	1,03194
2	Sumatera Utara	0,783901	2,113433
3	Sumatera Barat	5,828773	7,848981
4	Riau	7,549762	9,501284
5	Jambi	3,7674	5,800965
6	Sumatera Selatan	13,24666	15,28431
7	Bengkulu	0	2,077619
8	Lampung	10,57372	12,59962
9	Kep. Bangka Belitung	4,916849	6,801889
10	Kep. Riau	4,197273	6,150659
11	DKI Jakarta	28,10341	30,00851
12	Jawa Barat	4,709246	6,709538
...
...
32	Maluku Utara	3,6755	1,800722
33	Papua Barat	3,37433	1,435479
34	Papua	1,935924	2,637973
Jumlah		153,6378	171,9492
TotalCost		325,586951	

3. Hitung Total Simpangan (S)

Setelah didapatkan nilai jarak antara iterasi ke-1 dan iterasi ke-2, hitung total simpangan (S) dengan menghitung nilai total *distance* baru – total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan *k* objek baru sebagai *medoid*.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$= 325,586951 - 344,9809702$$

$$= -19,394$$

4. Ulangi langkah sebelumnya, tukar nilai objek dengan menentukan medoid baru. Hitung kembali jarak dari setiap objek pada iterasi ke-3 dengan medoid baru menggunakan persamaan (1). Setelah didapatkan nilai jarak iterasi ke-3, hitung kembali total simpangan (S) dengan mencari selisih dari nilai total *cost* baru- nilai total *cost* lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

$$= 634,2401363 - 325,586951$$

$$= 308,6532$$

Karena nilai $S > 0$ maka proses iterasi dihentikan. Sehingga diperoleh anggota tiap *cluster* pada tabel 6 sebagai berikut :

Tabel 6. Hasil Pengklasteran Dengan K-Medoids Clustering

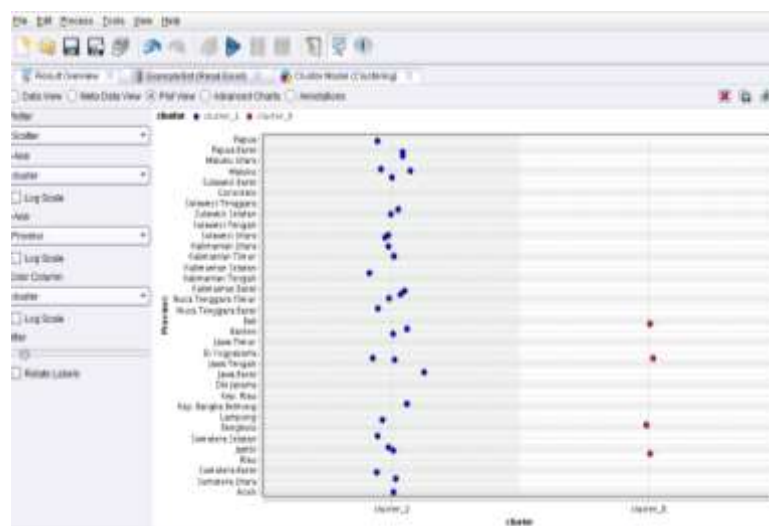
No	Provinsi	Jarak ke Medoid			Cluster yang Diikuti
		C1	C2	Terdekat	
1	Aceh	3,346311	17,16174	3,346311	1
2	Sumatera Utara	2,632204	15,21287	2,632204	1
3	Sumatera Barat	5,905861	9,115816	5,905861	1
4	Riau	8,046422	8,988237	8,046422	1
5	Jambi	4,056267	11,1679	4,056267	1
6	Sumatera Selatan	13,41325	3,304209	3,304209	2
7	Bengkulu	1,935924	14,90331	1,935924	1
8	Lampung	10,54511	4,346964	4,346964	2

9	Kep. Bangka Belitung	5,310951	10,97971	5,310951	1
10	Kep. Riau	4,834966	11,42677	4,834966	1
11	DKI Jakarta	27,99564	14,89663	14,89663	2
12	Jawa Barat	4,743016	10,20899	4,743016	1
13	Jawa Tengah	2,254107	16,21025	2,254107	1
...
...
32	Maluku Utara	4,349471	18,53072	4,349471	1
33	Papua Barat	3,756341	18,13862	3,756341	1
34	Papua	0	14,79003	0	1
Jumlah		171,1811	463,0591		
Totalcost		634,2401363			

Selanjutnya akan dilakukan pengolahan data menggunakan tools rapidminer untuk validasi hasil perhitungan manual dengan menggunakan aplikasi dengan *K-Medoids Clustering*. Pada tahap ini akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan aplikasi *Rapid Miner 5.3*.. Hasil akhir dapat dilihat pada gambar berikut :

Gambar 1. Hasil Akhir *Clustering*

Sehingga berdasarkan data hasil *clustering* pada gambar 4.10. didapatkan grafik hasil dari *Rapid Miner 5.3*. Dapat dilihat pada gambar 2. berikut :



Gambar 2. Grafik Hasil *Clustering* Persentase Kejahatan

Grafik menjelaskan hasil dari pengujian data persentase kejahatan pencurian dengan kekerasan di Indonesia menghasilkan 2 cluster yakni cluster tinggi dan cluster rendah. Dari hasil perhitungan *k-medoids* diperoleh sebanyak 4 provinsi sebagai cluster tinggi dan 30 Provinsi sebagai cluster rendah. Pada data yang telah dianalisa disimpulkan bahwa data yang digunakan adalah *valid*. Ini dibuktikan dengan hasil akhir perhitungan manual dengan hasil akhir dari aplikasi *Rapid Miner 5.3*. dapat menampilkan hasil akhir yang sama. Berikut adalah tampilan *cluster model* berupa teks yang ada dalam aplikasi *Rapid Miner 5.3*. dimana tampilan ini adalah penjelasan dari cluster yang terbentuk:



Gambar 3. Cluster Model Data Persentase Kejahatan

4. KESIMPULAN

Berdasarkan pembahasan dan hasil penelitian ini, maka di peroleh kesimpulan sebagai berikut:

1. Metode K-medoids Clustering dapat di implementasikan pada Penerapan Algoritma K-Medoids Dalam Mengelompokkan Tingkat Kasus Kejahatan di Setiap Provinsi.
2. Pada halaman K-medoids iterasi pertama sampai iterasi ketiga, hasilnya yaitu dari 34 data provinsi terdapat 2 provinsi masuk dalam cluster pertama (C1) yang dapat diartikan masuk dalam jumlah tingkat kejahatan yang paling tinggi..

REFERENCES

- [1] D. F. Pramesti, M. T. Furqon, And C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 1, No. 9, Pp. 723–732, 2017.
- [2] Z. Mustofa And I. S. Suasana, "Algoritma Clustering K-Medoids Pada E-Government Bidang Information And Communication Technology Dalam Penentuan Status Edgi," Vol. 9, Pp. 1–10, 2018.
- [3] S. Defiyanti, M. Jajuli, And N. R. W, "Optimalisasi K - Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion," *Teknosi*, Vol. 03, No. 01, Pp. 211–218, 2017.
- [4] H. Zayuka, S. M. Nasution, And Y. Purwanto, "Perancangan Dan Analisis Clustering Data Menggunakan Metode K-Medoids Untuk Berita Berbahasa Inggris," In *E-Proceeding Of Engineering*, 2017, Vol. 4, No. 2, Pp. 2182–2190.
- [5] D. Listiyanti, Y. A. Syahbana, And S. R. Henim, "Perancangan Dan Implementasi Aplikasi Android Penentu Salient Area Pada Video Dengan Algoritma," In *Annual Research Seminar 2016*, 2016, Vol. 2, No. 1, Pp. 96–101.