

KOMPARASI ALGORITMA K-MEANS DENGAN K-MEDOIDS DALAM KLASTERISASI WILAYAH RAWAN BENCANA DI KABUPATEN SITUBONDO

Ganang Aji Pambudhi¹⁾, Ahmad Homaidi²⁾, Firman Santoso³⁾

^{1,2,3)} Jurusan Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy
Jl. KHR. Syamsul Arifin No.1-2, Sukorejo, Situbondo 68374, Jawa Timur, Indonesia
E-mail : ¹⁾ganankaji@gmail.com, ²⁾ahmadhomaidi@ibrahimiy.ac.id,
³⁾firman4bi@gmail.com

ABSTRAK

Data dari Badan Penanggulangan Bencana Daerah Kabupaten Situbondo menunjukkan bahwa banyak bencana terjadi di Kabupaten Situbondo, baik yang disebabkan oleh alam maupun non-alam. Pemerintah dapat menentukan prioritas penanggulangan bencana dengan menggunakan hasil dari pembagian wilayah berdasarkan tingkat kerawanan bencana. Dengan menetapkan prioritas, sumber daya yang terbatas dapat dialokasikan ke wilayah yang paling membutuhkan. Dengan menggunakan pemodelan clustering, proses data mining dapat menggunakan data kejadian bencana untuk mengelompokkan daerah yang rawan bencana. Tujuan dari penelitian ini adalah untuk membandingkan pemodelan clustering daerah rawan bencana menggunakan algoritma K-Means dan K-Medoids, yang didasarkan pada data kejadian bencana dari tahun 2019 hingga 2023. Penelitian ini dilakukan dengan melihat hasil cluster dari kedua algoritma ini, menggunakan dataset yang sama dan jumlah cluster yang sama, yaitu tiga cluster. Hasil penelitian menunjukkan bahwa algoritma K-Means memiliki nilai indeks Davies-Bouldin (DBI) yang lebih baik (0.853) daripada K-Medoids (1.388). Hasil ini menunjukkan bahwa algoritma K-Means lebih efektif dalam mengelompokkan daerah Kabupaten Situbondo yang rawan bencana berdasarkan data kejadian bencana.

Kata kunci : Bencana, Data Mining, K-Means, K-Medoids, Clustering, Davies-Bouldin Index

ABSTRACT

Regional Disaster Management Agency's data reported that many disaster events happened in Situbondo Regency, both caused by natural and non-natural factors. The government can determine priorities in disaster management by utilizing the results of grouping regions based on disaster vulnerability. Level. Setting priorities allows limited resources to be allocated to areas of greatest need. The data mining process can use disaster event data to group disaster-prone areas by using clustering modeling. Therefore, this research aims to compare the clustering modeling of disaster-prone areas using the K-Means and K-Medoids algorithms based on disaster event data from 2019 to 2023. This research was carried out by looking at the cluster results of these two algorithms, using the same dataset and number of clusters, namely three clusters. The research results showed that the K-Means algorithm had a better Davies-Bouldin index (DBI) value (0.853) than K-Medoids (1.388). These results indicated that the K-Means algorithm was more effective in classifying disaster-prone areas of Situbondo Regency based on disaster event data.

Keywords: Disaster, Data Mining, K-Means, K-Medoids, Clustering, Davies-Bouldin Index

1. PENDAHULUAN

1.1 Latar Belakang Penelitian

Kabupaten Situbondo adalah salah satu daerah di Jawa Timur yang berpotensi untuk terjadi bencana. Kabupaten Situbondo terdiri dari tujuh belas kecamatan., dimana masing-masing kecamatan memiliki karakteristik geografis yang berbeda-beda. Bagian utara wilayah kabupaten situbondo

merupakan perairan yang berbatasan langsung dengan selat madura. Tidak hanya garis Pantai yang membentang dari ujung barat sampai ujung timur, pada bagian selatan juga terdapat Kawasan pegunungan atau dataran tinggi. Tercatat banyak kerugian harta benda dan korban terdampak bencanayang diakibatkan oleh berbagai bencana alam dan non-alam.

Sesuai dengan Undang-Undang Nomor 24 tahun 2007, upaya penanggulangan bencana membutuhkan perencanaan yang matang, terarah, dan terpadu. Undang-undang tersebut mendefinisikan penyelenggaraan penanggulangan bencana sebagai serangkaian tindakan yang mencakup kebijakan terkait pembangunan yang dapat menyebabkan bencana, pencegahan bencana, tindakan darurat, dan rehabilitasi [1]. Beragamnya jenis bencana di Kabupaten Situbondo menjadi tantangan tersendiri untuk dapat menciptakan inovasi serta strategi mitigasi bencana. Data dan informasi kebencanaan akan bermanfaat dalam perencanaan strategis penanggulangan bencana sehingga dapat dijadikan bahan untuk menetapkan kebijakan.

Data mining merupakan proses untuk mengidentifikasi pola dan pengetahuan berharga dari data yang ada. [2]. Data dan informasi perlu diolah menjadi pengetahuan sehingga dapat di analisis lebih lanjut serta menambah nilai manfaat dari suatu data. Dengan mempertimbangkan karakteristik data kejadian bencana yang diperoleh dari Badan Penanggulangan Bencana Daerah Kabupaten Situbondo, metode data mining yang digunakan adalah clustering menggunakan algoritma K-Means dan K-Medoids. Daerah rawan bencana yang ada di Kabupaten Situbondo akan dikelompokkan dengan metode ini.

Penelitian ini diharapkan dapat membantu pihak terkait merumuskan kebijakan untuk mengurangi risiko bencana dengan menggunakan algoritma K-Means dan K-Medoids. Selain menerapkan proses data mining tujuan dari penelitian ini juga untuk membandingkan algoritma yang paling optimal untuk menghasilkan kluster atau kelompok wilayah rawan bencana. Melalui penelitian ini diharapkan dapat meningkatkan kewaspadaan dan kesiapsiagaan terhadap bencana.

1.2 Tinjauan Penelitian

Nanda dan rekan-rekannya dalam jurnal "MALCOM: Indonesian Journal of Machine Learning and Computer Science" melakukan penelitian untuk membandingkan penggunaan algoritma K-Means dan K-Medoids dalam pengelompokan data kemiskinan. Penelitian ini menunjukkan bahwa algoritma K-Means lebih efektif dalam pengelompokan dibandingkan K-Medoids. Mereka menemukan bahwa dengan Davies Bouldin Index (DBI) sebesar 0,041, pengelompokan terbaik terjadi saat $k=8$ untuk K-Means, sementara K-Medoids mencapai hasil terbaik saat $k=2$ dengan DBI 0,52 [3]. Penelitian selanjutnya juga dilakukan oleh Mala Nafilah, Nining Rahaningsih, Raditya Danar

Dana dalam jurnal "JATI (jurnal Mahasiswa dan Teknik Informatika)" dengan tujuan untuk membandingkan pengelompokan produk pertanian menggunakan algoritma K-Means dan K-medoids. Berdasarkan hasil perbandingan antara K-Means dan K-Medoids, ditemukan bahwa K-Means mencapai nilai DBI sebesar 0,368 dengan pembagian data menjadi 3 kluster, sedangkan K-Medoids mencapai nilai DBI sebesar 0,706 dengan pembagian data menjadi 8 kluster. Melihat hasil percobaan, dapat disimpulkan bahwa nilai optimal untuk K-Means adalah 0,368 dengan pengelompokan menjadi 3 kluster [4].

Kecocokan dan relevansi topik dengan perbandingan algoritma pengelompokan daerah rawan bencana K-Means dan K-Medoids ditunjukkan oleh kedua penelitian. Adapun hal penting yang disampaikan oleh kedua penelitian tersebut adalah terkait pemilihan algoritma menjadi sangat penting dalam memperoleh hasil data mining berupa cluster. Penentuan daerah rawan bencana pemilihan algoritma yang tepat sesuai dengan karakteristik dataset diharapkan dapat menghasilkan pengelompokan daerah rawan bencana berdasarkan 3 tingkat kerawanan yaitu rendah, sedang, dan tinggi. Hasil dari kedua penelitian yang sudah dilakukan sebelumnya menjadi dasar untuk melakukan penelitian yang hampir sama dengan dataset yang berbeda.

1.3 Landasan Teori

a. Data Mining

Data mining adalah gabungan yang terpadu antara pengetahuan bisnis dan analisis statistik yang diperoleh dari data. Proses ini menggunakan berbagai teknik seperti statistik, matematika, kecerdasan buatan, simulasi, dan machine learning untuk mengekstraksi serta mengidentifikasi informasi yang relevan dari berbagai basis data yang berskala besar. [5]. Salah satu tujuan dari data mining adalah mengelompokkan atau klusterisasi data ke dalam segmen-segmen yang lebih kecil berdasarkan karakteristik tertentu dan lebih homogen.

b. Clustering

Merupakan metode pengelompokan data yang bertujuan untuk menempatkan data, pengamatan, atau kelompok ke dalam kelas-kelas. Penggunaan Pemodelan clustering dapat diterapkan pada dataset yang tidak memiliki label atau klasifikasi [6].

c. K-Means

Penerapan algoritma K-Means adalah salah satu teknik yang bisa dipakai untuk melakukan pengelompokan data. K-Means membagi data ke

yang bersifat preventif yaitu mitigasi bencana. Dalam melaksanakan urusan pemerintahan bidang penanggulangan bencana, Badan Penanggulangan Bencana Daerah Kabupaten Situbondo menghadapi masalah karena sumber daya yang terbatas, terutama dalam hal anggaran. Mempertimbangkan adanya keterbatasan tersebut maka diperlukan penentuan prioritas daerah atau wilayah rawan bencana di Kabupaten Situbondo yang terdiri dari 17 Kecamatan. Berdasarkan permasalahan tersebut maka klasterisasi dapat bermanfaat untuk mengelompokan daerah rawan bencana, dimana kelompok atau klaster daerah rawan bencana tersebut dapat menjadi pengetahuan untuk menentukan prioritas penanggulangan bencana yang efektif dan efisien.

3.2 Data Understanding

Data Kebencanaan yang relevan dalam menentukan daerah rawan bencana adalah data historis kejadian bencana. Penelitian ini menggunakan data kejadian bencana kabupaten Situbondo dari tahun 2019 hingga 2023. Adapun data yang termuat antara lain nama kecamatan, jumlah kejadian, jenis bencana, serta tahun dengan jumlah 204 baris. Sebagian besar data yang disajikan bertipe numerik serta tidak ada yang menunjukkan kategorisasi dan klasifikasi. Data mentah tersebut masih belum dapat menyajikan informasi yang jelas terkait jumlah kejadian bencana berdasarkan wilayah yaitu kecamatan termasuk jenis bencana. Oleh karena itu dibuat tabel baru yang berisi rekapitulasi jumlah data kejadian bencana pada tahun 2019 sampai dengan tahun 2023 berdasarkan wilayah kecamatan dan jenis bencana. Berikut adalah tabel rekapitulasi kejadian bencana :

Tabel 1. Rekapitulasi Kejadian Bencana

Nama Kecamatan	a	b	c	d	e	f	g	h	i	j	K	L
Situbondo	0	1	0	27	0	11	0	0	0	23	0	0
Panji	0	0	0	20	0	14	0	0	0	19	0	0
..
Kapongan	0	1	5	12	0	3	0	0	0	13	0	0

*keterangan atribut

- a: Banjir
- b: Tanah Longsor
- c: Gelombang pasang dan Abrasi
- d: cuaca ekstrem
- e: kekeringan
- f: Kebakaran hutan dan lahan
- g: Gempabumi
- h: Tsunami

- i: Erupsi Gunung Api
- j: Kebakaran gedung dan pemukiman
- k: Gagal Teknologi
- l: Konflik Sosial

Selain itu juga dilakukan pemisahan data kecamatan untuk mengorganisasi data dan menghindari redundansi, sehingga terdapat tabel tambahan berupa data kecamatan. Berikut adalah tabel tabel kecamatan bencana :

Tabel 2. Data Kecamatan

id	Nama Kecamatan
1	Situbondo
2	Panji
..	..
17	Kapongan

3.3 Data Preparation

Pada tahapan ini dilakukan proses perbaikan dan penyesuaian terhadap data awal sampai siap untuk dijadikan dataset. Adapun proses pertama yang dilakukan adalah melakukan relasi antara tabel kecamatan dengan tabel rekapitulasi kejadian bencana. Selanjutnya melakukan seleksi terhadap atribut, dari total 14 atribut menjadi 13 atribut. Sehingga di peroleh dataset sesuai tabel berikut :

Tabel 3. Dataset

id	a	b	c	d	e	f	g	h	i	j	K	L
1	0	1	0	27	0	11	0	0	0	23	0	0
2	0	0	0	20	0	14	0	0	0	19	0	0
..
17	0	1	5	12	0	3	0	0	0	13	0	0

Berikut adalah penjelasan terhadap masing-masing atribut yang pada dataset :

Tabel 4. Penjelasan Atribusi

Atribut	Tipe Data	Deskripsi
id	integer	Nilai numerik untuk kode kecamatan
a	integer	Nilai numerik yang berisi jumlah kejadian bencana banjir
b	integer	Nilai numerik yang berisi jumlah kejadian tanah longsor

Atribut	Tipe Data	Deskripsi
c	integer	Nilai numerik yang berisi jumlah kejadian gelombang pasang dan abrasi
d	integer	Nilai numerik yang berisi jumlah kejadian cuaca ekstrem
e	integer	Nilai numerik yang berisi jumlah kejadian kekeringan
f	integer	Nilai numerik yang berisi jumlah kejadian kebakaran hutan dan lahan
g	integer	Nilai numerik yang berisi jumlah kejadian gempa bumi
h	integer	Nilai numerik yang berisi jumlah kejadian tsunami
i	integer	Nilai numerik yang berisi jumlah kejadian erupsi gunung api
j	integer	Nilai numerik yang berisi jumlah kejadian kebakaran gedung dan pemukiman
k	integer	Nilai numerik yang berisi jumlah kejadian gagal teknologi
l	integer	Nilai numerik yang berisi jumlah kejadian konflik sosial

model cluster terhadap dataset dapat dilihat pada tabel berikut :

Tabel 5. Klaster Algoritma K-Means

id	Klaster 0	Klaster 1	Klaster 2
1	1		
2	1		
3	1		
4		1	
5	1		
6			1
7		1	
8		1	
9	1		
10			1
11			1
12	1		
13			1
14	1		
15	1		
16	1		
17	1		
jumlah	10	3	4
total		17	

Berdasarkan tabel diatas dapat disimpulkan Penggunaan algoritma K-Means untuk mengelompokan wilayah rawan bencana berdasarkan jumlah kejadian per jenis bencana menghasilkan klaster tingkat rendah (klaster 0) yaitu kecamatan situbondo, panji, mangaran, jangkar, besuki, jatibanteng, bungatan, banyuglugur, asempagus, dan kapongan. Kemudian klaster tingkat sedang (klaster 1) berada di kecamatan banyuputih, panarukan, dan kendit. Klaster tingkat tinggi (klaster 2) berada di kecamatan arjasa, mlandingan, suboh, dan sumbermalang

Algoritma K-Medoids akan digunakan pada skema percobaan pemodelan data mining selanjutnya, dan algoritma ini juga dapat membantu dalam clustering. Dataset dan jumlah klaster yang ditentukan sama dengan jumlah percobaan sebelumnya yaitu 3 klaster. Tiga klaster yang dibuat dengan pemodelan adalah klaster 0 dengan sebelas item, klaster 1 dengan empat item, dan klaster 2 dengan dua item. Tabel berikut menunjukkan hasil penggunaan algoritma tersebut.

Tabel 6. Klaster Algoritma K-Medoids

id	Klaster 0	Klaster 1	Klaster 2
1	1		
2	1		
3	1		
4		1	

3.4 Modeling

Pada tahapan ini dilakukan pemodelan terhadap dataset sehingga nantinya akan menghasilkan model berupa klaster. Dalam menghasilkan model berupa klaster, algoritma K-Means dan K-Medoids digunakan bersama dengan alat data mining seperti RapidMiner.. Peneliti menetapkan jumlah klaster atau nilai k sebanyak 3 klaster. Peneliti menggunakan aplikasi RapidMiner untuk membantu proses data mining.

Skema percobaan pertama adalah dengan melakukan pemodelan menggunakan algoritma K-Means. Langkah awal yang dilakukan adalah dengan menetapkan angka klaster, yaitu jumlah klaster yang ingin dihasilkan. Sebagaimana jumlah nilai klaster yang sudah ditetapkan oleh peneliti yaitu 3 klaster. Tentu saja Hasil pemodelan menunjukkan adanya 3 klaster, dengan klaster 0 memiliki 10 item, klaster 1 memiliki 3 item, dan klaster 2 memiliki 4 item. Hasil

id	Klaster 0	Klaster 1	Klaster 2
5	1		
6	1		
7		1	
8		1	
9	1		
10	1		
11			1
12	1		
13			1
14		1	
15	1		
16	1		
17	1		
jumlah	11	4	2
total		17	

Berdasarkan hasil sebaran klaster di setiap data maka wilayah yang termasuk tingkat rendah (klaster 0) meliputi kecamatan situbondo, panji, mangaran, jangkar, arjasa, besuki, mlandingan, jatibanteng, banyuglugur, asempagus, dan kapongan. Selanjutnya untuk wilayah yang berada di tingkat sedang (klaster 1) meliputi kecamatan banyuputih, panarukan, kendit, dan bungatan. Wilayah dengan tingkat tinggi (klaster 2) berada di wilayah sumber malang dan suboh.

3.5 Evaluation

Setelah percobaan pemodelan clustering telah dilaksanakan maka tahapan selanjutnya adalah mengevaluasi hasil klaster yang didapatkan. Evaluasi dilakukan dengan cara membandingkan penggunaan algoritma K-means dengan algoritma K-Medoids. Sebelum membandingkan kedua algoritma tersebut terlebih dahulu mengecek kinerja 3 klaster terhadap masing-masing algoritma menggunakan Davies-Bouldin index (DBI). Pengujian DBI masing-masing algoritma dilakukan menggunakan aplikasi RapidMiner dengan memasukan operator cluster distance performance. Berikut tampilan hasil pengujian terhadap algoritma K-Means :

```

PerformanceVector
PerformanceVector:
Avg. within centroid distance: -186.269
Avg. within centroid distance_cluster_0: -144.790
Avg. within centroid distance_cluster_1: -196.889
Avg. within centroid distance_cluster_2: -282.000
Davies Bouldin: -0.853
    
```

Gambar 2. Hasil Pengujian DBI K-Means

Pengujian terhadap algoritma K-Medoids dapat dilihat pada gambar berikut :

```

PerformanceVector
PerformanceVector:
Avg. within centroid distance: -372.059
Avg. within centroid distance_cluster_0: -265.727
Avg. within centroid distance_cluster_1: -543.000
Avg. within centroid distance_cluster_2: -615.000
Davies Bouldin: -1.388
    
```

Gambar 3. Hasil Pengujian DBI K-Medoids

Perbandingan terhadap hasil pengujian kedua algoritma tersebut dapat dilihat pada tabel berikut :

Tabel 7. Hasil Perbandingan Algoritma Pemodelan Clustering

No	Algoritma	Jumlah k	Nilai DBI
1	K-Means	3	0.853
2	K-Medoids	3	1.388

Hasil perbandingan menunjukkan bahwa nilai DBI yang diperoleh dari clustering dengan algoritma K-Means lebih unggul dibandingkan dengan nilai DBI yang diperoleh dari clustering dengan K-Medoids. Hal ini terlihat dari nilai DBI algoritma K-Means yang lebih rendah yaitu 0,853 untuk jumlah cluster dan dataset yang sama.

3.6 Deployment

Setelah melakukan evaluasi terhadap hasil analisis data mining yaitu klaster, maka selanjutnya adalah memperkenalkan model atau hasil analisis tersebut melalui penyusunan laporan dan dokumentasi. Laporan hasil analisis diharapkan dapat membantu dalam pengambilan keputusan dalam aspek kebencanaan.

4. KESIMPULAN DAN SARAN

Pengelompokan daerah rawan bencana sangat penting untuk membantu pihak berwenang dalam menentukan prioritas penanggulangan bencana, merencanakan strategi mitigasi, dan meningkatkan kesadaran masyarakat terhadap risiko bencana. Berdasarkan nilai Davies-Bouldin Index (DBI), algoritma K-Means mengungguli K-Medoids dalam klasifikasi wilayah rawan bencana. Temuan pengelompokan ini dapat digunakan untuk memperbaiki kebijakan dan prosedur mitigasi bencana serta meningkatkan kesiapsiagaan masyarakat di daerah rawan bencana. Studi ini dapat menjadi landasan bagi penelitian selanjutnya mengenai algoritma atau metode tambahan yang dapat digabungkan sehingga dapat meningkatkan

presisi dan efisiensi pengelompokan daerah rawan bencana.

Tree Dengan Algoritma Cart Untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang,” *J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, pp. 367–380, 2022, [Online]. Available: <http://dx.doi.org/10.31941/delta.v10i2.2172>

DAFTAR PUSTAKA

- [1] Pemerintah Pusat, *Undang-undang (UU) Nomor 24 Tahun 2007 tentang Penanggulangan Bencana*. 2007.
- [2] Siti Nurajizah, “Analisa Transaksi Penjualan Obat Menggunakan Algoritma Apriori,” *INOVTEK Polbeng - Seri Inform.*, vol. 4, no. 1, pp. 35–44, 2019.
- [3] N. T. Luchia, H. Handayani, F. S. Hamdi, D. Erlangga, and S. F. Octavia, “Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 35–41, 2022, doi: 10.57152/malcom.v2i2.422.
- [4] M. Nafilah, N. Rahaningsih, and R. Danar Dana, “Perbandingan Algoritma K-Means Dan K-Medoids Untuk Pengelompokan Hasil Pertanian Di Kabupaten Cirebon,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 223–229, 2024, doi: 10.36040/jati.v8i1.8326.
- [5] A. Irfan, *Implementasi Data Mining Dalam Pemeliharaan Jalan*. CV. Amerta Media, 2021.
- [6] P. Putriana, N. Suarna, and W. Prihartono, “Analisis Clustering Prestasi Atlet Pada Berbagai Cabang Olahraga Menggunakan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3435–3442, 2024, doi: 10.36040/jati.v7i6.8211.
- [7] M. Murdiaty, A. Angela, and C. Sylvia, “Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means,” *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 744, 2020, doi: 10.30865/mib.v4i3.2213.
- [8] Dita, A. Munawwaroh, Arum, and H. Primandari, “Implementasi Crisp-Dm Model Menggunakan Metode Decision