

PENGELOMPOKAN DAERAH RAWAN BENCANA ALAM DI JAWA BARAT MENGUNAKAN ALGORITMA FUZZY C-MEANS

Ide Nabilla Audy, Tesa Nur Padilah, Betha Nurina Sari

Program Studi Informatika S1, Fakultas Ilmu Komputer

Universitas Singaperbangsa Karawang, Jalan HS. Ronggo Waluyo, Karawang, Indonesia

ide.nabilla19023@student.unsika.ac.id

ABSTRAK

Bencana alam adalah kejadian berbahaya yang diakibatkan oleh berbagai fenomena alam yang diakibatkan oleh alam. Jawa Barat adalah salah satu provinsi yang memiliki tingkat kejadian bencana alam tertinggi dan bisa juga disebut daerah yang rawan bencana alam. Untuk menanggulangi bencana alam dibutuhkan relawan yang banyak pada daerah yang termasuk rawan bencana. Maka dari itu, untuk mengetahui jumlah relawan yang dibutuhkan, perlunya informasi mengenai daerah mana saja yang termasuk daerah dengan tingkat bencana alam yang tinggi. Tujuan dari penelitian ini untuk mengelompokkan kabupaten/kota yang rawan bencana alam, sehingga dapat menjadi fokus bagi pemerintah dalam penanggulangan bencana alam. Dengan metode *Clustering Fuzzy C-Means* dipakai untuk mengelompokkan kabupaten/kota yang rawan dan tidak rawan. Hasil *clustering* yang paling optimal yaitu pada $c = 2$ dengan hasil evaluasi memiliki struktur yang baik yang dilihat dari hasil evaluasi dengan metode *Silhouette Coefficient*, *Partition Entropy (PE)*, dan *Partition Coefficient (PC)*.

Kata kunci: Data Mining, Clustering, Fuzzy C-Means, Bencana Alam

1. PENDAHULUAN

Bencana ialah peristiwa mengancam dan dapat mengganggu kehidupan manusia, penyebabnya bisa dari faktor alam, bukan faktor alam dan juga faktor manusia itu sendiri. Bencana alam adalah kejadian berbahaya yang diakibatkan oleh berbagai fenomena alam yang diakibatkan oleh alam. Setiap jenis kejadian bencana alam tersebut mempunyai bahaya dan dampak yang berbeda-beda tergantung pada penyebab terjadinya bencana alam tersebut [1].

Berdasarkan data dari Data Informasi Bencana Indonesia (DIBI), pada tahun 2017 sampai dengan tahun 2022 ada 18.861 bencana yang terjadi di Indonesia, dan jumlah kejadian tertinggi ada pada pulau Jawa setiap tahunnya. Dari hasil penelitian [2] provinsi Jawa Barat ialah salah satu provinsi yang memiliki tingkat kejadian bencana alam tertinggi dan bisa juga disebut daerah yang rawan bencana alam.

Di antara 11 bencana alam yang terjadi, terdapat 3 bencana alam besar yaitu tanah longsor, angin puting beliung dan banjir merupakan hal yang biasa terjadi di provinsi Jawa Barat. 3 bencana tersebut terjadi setiap tahunnya dari tahun 2017 sampai 2022.

Data mining ialah model yang dilakukan pada penelitian ini dengan metode *clustering*. *Clustering* adalah proses pengelompokan data dimana data yang sama pada satu kelompok akan lebih besar dari data yang sama pada kelompok lain [3].

Teknik yang digunakan untuk pengelompokan daerah rawan bencana alam pada penelitian ini yaitu menggunakan algoritma *Fuzzy C-Means* dikarenakan pada penelitian sebelumnya yaitu penelitian [4], mendapatkan hasil bahwa dari 2 algoritma yang digunakan, hasil yang lebih optimal didapatkan dengan menggunakan algoritma *Fuzzy C-Means*.

Untuk mewujudkan masyarakat yang memiliki pemahaman dalam penanggulangan bencana, diperlukannya relawan untuk membantu dan

memberikan informasi terkait upaya penanggulangan bencana. Relawan tersebut diberi nama TAGANA atau Taruna Siaga Bencana terdiri dari masyarakat yang mempunyai kepedulian tinggi terhadap penanggulangan bencana. Taruna Siaga Bencana (TAGANA) sudah tersebar di seluruh kota di Indonesia. Berdasarkan dataset dari Open Data Jabar, jumlah anggota TAGANA di Jawa Barat pada tahun 2021 berjumlah 1.791 orang.

Seiring berjalannya waktu, bencana alam semakin sering terjadi di beberapa provinsi di Indonesia, khususnya di Jawa Barat, dan pastinya membutuhkan lebih banyak anggota TAGANA. Maka dari itu, untuk mengetahui jumlah anggota TAGANA yang dibutuhkan, perlunya informasi mengenai daerah mana saja yang termasuk daerah dengan tingkat bencana alam yang tinggi.

2. TINJAUAN PUSTAKA

2.1. Bencana Alam

Bencana ialah peristiwa mengancam dan dapat mengganggu kehidupan manusia, penyebabnya bisa dari faktor alam, bukan faktor alam dan juga faktor manusia itu sendiri. Bencana alam merupakan kejadian berbahaya yang diakibatkan oleh berbagai fenomena alam yang diakibatkan oleh alam. Setiap jenis kejadian bencana alam tersebut mempunyai bahaya dan dampak yang berbeda-beda tergantung pada penyebab terjadinya bencana alam tersebut [1].

2.2. Data Mining

Data mining merupakan proses analisis data yang dirancang agar memberikan perspektif berbeda dan dapat diringkas menjadi informasi yang bermanfaat. Definisi sederhana dari *data mining* adalah mengekstraksi sebuah pengetahuan yang penting dan menarik dari data dalam database yang besar [5]. Terdapat beberapa teknik pada data mining yakni

estimasi, *forecasting*, *classification*, *clustering* dan *association*.

2.3. Clustering

Clustering adalah proses pengelompokan kelas-kelas yang memiliki bentuk yang sama. *Clustering* terjadi pada awal proses *data mining* selama fase analisis [6]. Algoritma yang ada pada teknik *clustering* yaitu *K-Means*, *K-Medoids*, *Fuzzy C-Means*, dan lain-lain.

2.4. Fuzzy C-Means

Fuzzy C-Means (FCM) adalah algoritma *clustering* yang termasuk dalam metode *Hard K-Means* [7]. *Fuzzy C-Means (FCM)* merupakan suatu cara mendapatkan nilai sebenarnya dari setiap objek berdasarkan derajat keanggotaannya dan menugaskan objek tersebut ke dalam satu atau lebih *cluster* [4]. Kelebihan dari algoritma ini yaitu terletak pada penempatan pusat *cluster* nya lebih tepat daripada algoritma lain. Akan tetapi, kelemahan dari algoritma ini yaitu terletak pada penentuan *cluster* optimalnya, maka dari itu diperlukannya tahapan validasi.

2.5. Silhouette Coefficient

Silhouette Coefficient merupakan metode yang ada pada teknik *clustering* untuk menentukan validasi. Rentang nilai *Silhouette Coefficient* bervariasi dari -1 hingga 1. Semakin mendekati 1, jumlah *cluster* akan menjadi lebih baik. Sebaliknya, apabila dekat dengan -1 maka *cluster* tersebut semakin buruk [4]. Berikut adalah rentang nilai dan kategori pada *Silhouette Coefficient*.

Tabel 1. Rentang dan Kategori *Silhouette Coefficient*

Rentang	Kategori
$\geq 0.71 \leq 1.00$	Kuat
$\geq 0.51 \leq 0.70$	Baik
$\geq 0.26 \leq 0.50$	Lemah
≤ 0.25	Buruk

2.6. Partition Entropy (PE)

Partition Entropy (PE) merupakan metode yang ada pada teknik *clustering* untuk menentukan validasi. Faktanya, metode *Partition Entropy (PE)* adalah metode pertama kali yang terkait dengan algoritma *Fuzzy C-Means*. Jumlah *cluster* optimal terjadi ketika nilainya mendekati 0. Semakin kecil nilai yang didapat maka kualitas *cluster* tersebut akan semakin baik [8]. Berikut adalah persamaan dari *Partition Entropy*.

$$PE = \frac{1}{N} \left(\sum_{i=1}^N \sum_{j=1}^k U_{ij} \times \log U_{ij} \right)$$

2.7. Partition Coefficient (PC)

Partition Coefficient (PC) merupakan metode yang ada pada teknik *clustering* untuk menentukan validasi. Rentang nilai pada metode ini yaitu dari 0 hingga 1, semakin besar nilai yang didapat maka

kualitas *cluster* tersebut semakin baik [8]. Berikut adalah persamaan dari *Partition Coefficient*.

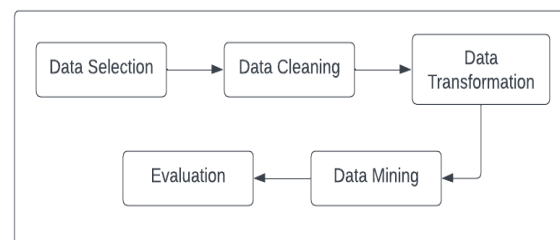
$$PC = \frac{1}{N} \left(\sum_{i=1}^N \sum_{j=1}^k U_{ij}^2 \right)$$

2.8. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan metode *data mining* yang mempunyai tujuan mendapatkan pengetahuan yang berguna untuk pengambilan keputusan [9]. Manfaat dari metode ini yaitu pada bidang bisnis dapat membantu memprediksi tren pasar dimasa depan, meningkatkan efisiensi kerja, peningkatan layanan konsumen dan lain-lain.

3. METODE PENELITIAN

Knowledge Discovery in Database (KDD) ialah metode *data mining* yang akan digunakan pada penelitian ini. Metode ini dibagi menjadi 5 tahap seperti yang dijelaskan pada gambar di bawah ini.



Gambar 1. Metode Penelitian

3.1. Data Selection

Proses yang dilakukan pada tahapan ini yaitu melakukan pengumpulan dan pemilihan data. Dataset pada penelitian ini diperoleh dari *situs website* Open Data Jabar dan *website* resmi DIBI Badan Nasional Penanggulangan Bencana. Dari data yang telah didapat, dipilih sebanyak 162 data yang didapatkan dari 27 kabupaten/kota di provinsi Jawa Barat per tahun 2017 sampai tahun 2022.

3.2. Data Cleaning

Proses yang dilakukan pada tahapan ini yaitu melakukan pembersihan data. Pada tahap ini akan dilakukan 2 tahap pengecekan yaitu pengecekan *missing value* dan pengecekan *outlier*. Setelah pengecekan dilakukan, *dataset* yang terdapat *missing value* atau *outlier* akan ditangani dengan cara *winsorize mean* yakni mengganti nilai *outlier* dengan nilai *mean*.

3.3. Data Transformation

Proses yang dilakukan pada tahapan ini yaitu melakukan perubahan tipe data dari data *non numeric* menjadi data *numeric* yang bertujuan untuk mendapatkan pola yang sesuai dan memudahkan pada proses *data mining*.

3.4. Data Mining

Proses yang dilakukan pada tahapan ini yaitu melakukan pengolahan data menggunakan metode

clustering dengan algoritma *fuzzy c-means* dan menggunakan tools *RStudio*.

3.5. Evaluation

Pada tahapan ini dilakukan proses evaluasi. Metode evaluasi yang digunakan pada penelitian ini adalah metode *Silhouette Coefficient*, *Partition Entropy (PE)*, dan *Partition Coefficient (PC)*.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Hasil dari seleksi data tersebut mendapatkan 10 atribut yang digunakan, yaitu nama kabupaten/ kota, jumlah kejadian bencana tanah longsor, jumlah kejadian bencana angin puting beliung, jumlah kejadian bencana banjir, jumlah korban bencana tanah longsor, jumlah korban bencana angin puting beliung, jumlah korban bencana banjir, jumlah kerusakan bencana tanah longsor, jumlah kerusakan bencana angin puting beliung, dan jumlah kerusakan bencana banjir.

4.2. Data Cleaning

Pada tahap ini akan dilakukan 2 tahap pengecekan yaitu pengecekan *missing value* dan pengecekan *outlier*. Setelah pengecekan dilakukan, dataset yang terdapat *missing value* atau *outlier* akan ditangani.

1. Pengecekan Missing Value

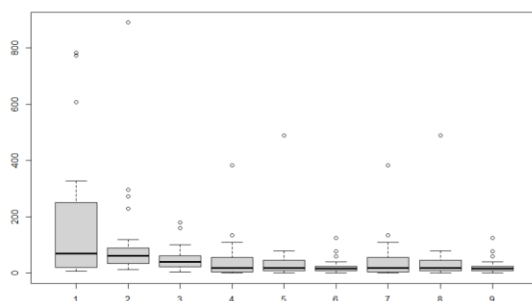
Hasil *missing value* dari setiap variabel yaitu bernilai 0 yang berarti tidak terdapat *missing value* pada variabel tersebut yang ditunjukkan pada Gambar 2.

```
> colSums(is.na(Bencana_Alam_Gabungan))
Nama_Kabupaten_Kota      Jumlah_Tanah_Longsor      Jumlah_Puting_Beliung
Jumlah_Banjir      Korban_Tanah_Longsor      Korban_Puting_Beliung
Korban_Banjir      Kerusakan_Tanah_Longsor      Kerusakan_Puting_Beliung
Kerusakan_Banjir
```

Gambar 2. Hasil Pengecekan Missing Value

2. Pengecekan Outlier

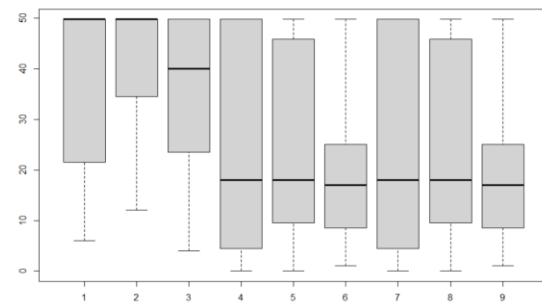
Hasil pengecekan *outlier* dari setiap variabel yaitu menunjukkan bahwa terdapat nilai ekstrim pada semua variabel yang dilihat pada Gambar 3.



Gambar 3. Hasil Pengecekan Outlier

Dilihat dari hasil *outlier* yang didapat, maka akan dilakukan penanganan pada *outlier* dengan cara *winsorize mean*. *Winsorize mean* ini dilakukan dengan mengganti nilai *outlier* menjadi *mean*. Hasil dari

penanganan *outlier* menggunakan cara *winsorize mean* dapat dilihat pada Gambar 4.



Gambar 4. Hasil Penanganan Outlier

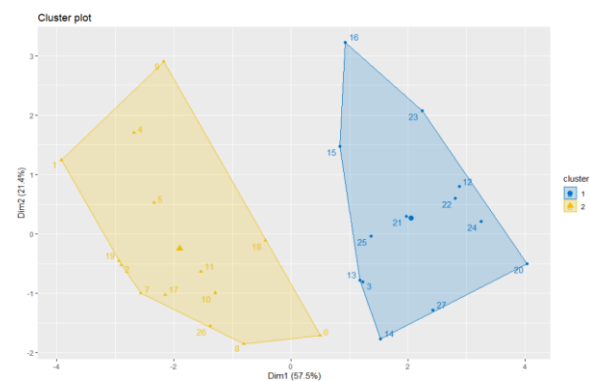
Dapat dilihat bahwa hasil setelah penanganan *outlier* dilakukan, pada semua variabel yang sebelumnya terdapat *outlier* menjadi bersih dari *outlier*.

4.3. Data Transformation

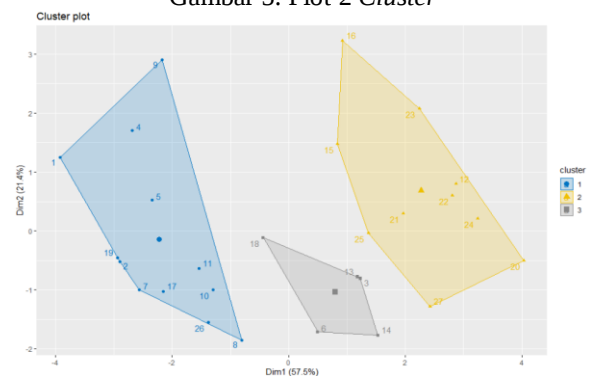
Setelah pembersihan data dilakukan pada tahap sebelumnya, maka selanjutnya akan dilakukan tahap transformasi data/ *data tranformation*. Dari data yang telah tersedia, tidak perlu melakukan transformasi data karena data yang akan digunakan telah berbentuk *numeric*.

4.4. Data Mining

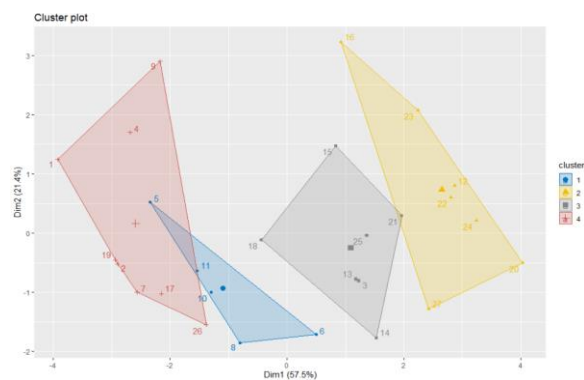
Hasil dari *clustering* menggunakan algoritma *fuzzy c-means* dengan $c = 2$, $c = 3$, dan $c = 4$ yaitu sebagai berikut.



Gambar 5. Plot 2 Cluster



Gambar 6. Plot 3 Cluster



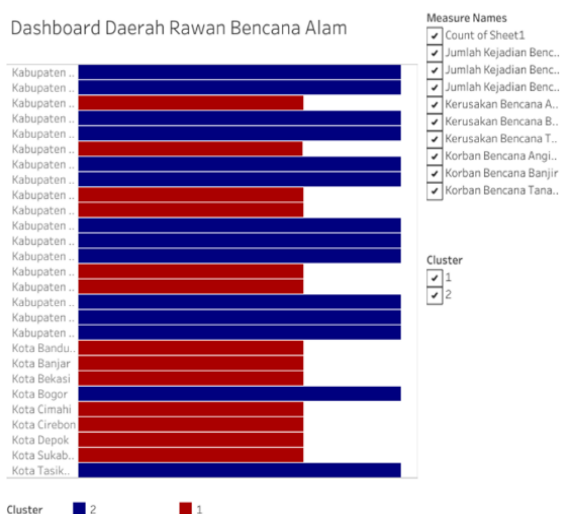
Gambar 7. Plot 4 Cluster

4.5. Evaluation

Hasil *evaluation* menggunakan metode *Silhouette Coefficient*, *Partition Entropy (PE)*, dan *Partition Coefficient (PC)* yaitu sebagai berikut.

Tabel 1. Hasil *Evaluation*

C	<i>Silhouette Coefficient</i>	<i>Partition Entropy (PE)</i>	<i>Partition Coefficient (PC)</i>
2	0.68	0.40	0.74
3	0.53	0.74	0.56
4	0.44	0.87	0.53

Gambar 8. Visualisasi *Dashboard*

Evaluasi yang dilakukan yaitu menggunakan 3 metode antara lain yaitu metode *Silhouette Coefficient* dengan nilai 0,68 yang menunjukkan bahwa hasil *clustering* memiliki struktur baik, metode *Partition Entropy (PE)* dengan nilai 0,40, dan metode *Partition Coefficient* dengan nilai 0,74.

Berdasarkan hasil evaluasi yang dilakukan *cluster* yang paling optimal yaitu pada 2 *cluster*, maka dari itu kabupaten/kota di Jawa Barat akan dikelompokkan menjadi 2 *cluster* untuk menentukan kabupaten/kota mana yang berkategori rawan bencana alam dan tidak rawan bencana alam.

Berikut adalah hasil visualisasi *clustering* yang berupa *dashboard* menggunakan tools *Tableau Public*. Dari 27 kabupaten/kota dilakukan pengelompokan menjadi 2 *cluster* yakni pada *cluster* 1 ditandai dengan warna merah dan *cluster* 2 ditandai dengan biru.

5. KESIMPULAN DAN SARAN

Dari penelitian yang dilakukan dapat disimpulkan bahwa *clustering* yang paling optimal yaitu pada $c = 2$ yang didapat dari hasil evaluasi menggunakan metode *Silhouette Coefficient*, *Partition Entropy (PE)*, dan *Partition Coefficient (PC)*.

Hasil *clustering* daerah rawan bencana alam menggunakan algoritma *Fuzzy C-Means* menunjukkan hasil kabupaten/kota yang rawan bencana alam yakni pada tahun 2017-2022 yaitu terdapat sebanyak 14 kabupaten/kota yaitu Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Bandung, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Ciamis, Kabupaten Kuningan, Kabupaten Cirebon, Kabupaten Majalengka, Kabupaten Sumedang, Kabupaten Bandung Barat, Kabupaten Pangandaran, Kota Bogor, dan Kota Tasikmalaya.

Dengan dibuatnya penelitian ini maka diharapkan dapat menjadi acuan bagi penelitian selanjutnya. Adapun hal yang peneliti ingin sarankan untuk penelitian selanjutnya adalah dapat menggunakan algoritma, variabel dan metode evaluasi lain yang lebih baik dan lebih beragam. Selain itu, diharapkan pada penelitian selanjutnya menghasilkan lokasi daerah rawan bencana secara detail sampai ke titik lokasi terjadinya.

DAFTAR PUSTAKA

- [1] I. N. Setiawan, D. Krismawati, S. Pramana, dan E. Tanur, "Klasterisasi wilayah rentan bencana alam berupa gerakan tanah dan gempa bumi di Indonesia," *Semin. Nas. Off. Stat.*, (1), 669–676, 2022, doi: 10.34123/semnasoffstat.v2022i1.1538.
- [2] M. S. Yana, L. Setiawan, E. M. Ulfa, dan A. Rusyana, "Penerapan metode k-means dalam pengelompokan wilayah menurut intensitas kejadian bencana alam di Indonesia tahun 2013-2018," *J. Data Anal.*, 1(2), 93–102, 2018, doi: 10.24815/jda.v1i2.12584.
- [3] M. G. Sadewo, A. P. Windarto, dan A. Wanto, "Penerapan algoritma clustering dalam mengelompokkan banyaknya desa/kelurahan menurut upaya antisipasi/ mitigasi bencana alam menurut provinsi dengan k-means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, 2(1), 311–319, 2018, doi: 10.30865/komik.v2i1.943.
- [4] H. Firdaus dan A. Sofro, "Analisa cluster menggunakan k-means dan fuzzy c-means dalam pengelompokan provinsi menurut data intensitas bencana alam di Indonesia tahun 2017-2021," *MATHunesa J. Ilm. Mat.*, 10(1), 50–60, 2022, doi: 10.26740/mathunesa.v10n1.p50-60.
- [5] E. Supriatna, "Information System Journal,"

- INTERNAL-Information syatem J.*, 3(2), 51–64, 2020.
- [6] E. Tasia dan M. Afdal, “Perbandingan algoritma k-means dan k-medoids untuk clustering daerah rawan banjir di kabupaten rokan hilir,” 3(1), 65–73, 2023.
- [7] C. Ifadah *et al.*, “Clustering daerah banjir di jawa timur dengan,” 14(2), 99–104, 2022.
- [8] A. Ramadhan dan R. Handinata, “Implementasi algoritma fuzzy c-means untuk pengelompokan wilayah bencana banjir,” 12, 2579–5406, 2019, [Daring]. Tersedia pada: www.dibi.bnpb.-.
- [9] Y. Mardi, “Data mining : klasifikasi menggunakan algoritma c4.5,” *Edik Inform.*, 2(2), 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.