

# CLUSTERING DAERAH BANJIR DI JAWA TIMUR DENGAN ALGORITMA FUZZY C-MEANS

Corii Ifadah,<sup>1</sup> Cristanti Dwi Ratnasri<sup>2</sup>, Dian Candra Rini Novitasari<sup>3</sup>

<sup>1,2,3</sup>Program Studi Matematika, Fakultas Sains & Teknologi, UIN Sunan Ampel Surabaya

e-mail: <sup>1</sup> corii1829@gmail.com, <sup>2</sup> cristantidwi5@gmail.com, <sup>3</sup> diancrini@uinsby.ac.id

## ABSTRAK

Bencana alam seperti banjir merupakan peristiwa yang dapat terjadi setiap waktu untuk dapat mengurangi resiko perlu diadakan mitigasi bencana. Namun pada tiap daerah memiliki karakteristik bencana banjir yang berbeda. Penelitian ini dimaksud untuk melakukan clustering kejadian banjir yang terjadi di Jawa Timur. Proses pengelompokan dilakukan dengan menggunakan metode Fuzzy C Means. Variabel yang digunakan pada penelitian ini yaitu jumlah kejadian, korban, dan kerusakan. Proses pengklasteran dan kejadian bencana dilakukan terhadap wilayah kabupaten/kota yang ada di Jawa Timur selama 4 tahun terakhir. Untuk pengklasteran ditentukan parameter dari  $c=3$ ,  $w=2$ ,  $MaxIter=100$ , dan  $\epsilon=0,0001$ . Diperoleh hasil algoritma Fuzzy C-means cluster 1: Pasuruan, cluster 2: Bojonegoro, Gresik, Jember, Jombang, Lamongan, cluster 3: Banyuwangi, Batu, Blitar, Bondowoso, Kediri, Lumajang, Madiun, Magetan, Malang, Mojokerto, Nganjuk, Ngawi, Pacitan, Pamekasan, Ponorogo, Probolinggo, Sampang, Sidoarjo, Situbondo, Sumenep, Surabaya, Trenggalek, Tuban, Tulungagung. Validasi hasil cluster dilakukan dengan Silhouette Coefficient diperoleh nilai sebesar 0,84377 nilai ini masuk dalam kriteria pengukuran yang kuat.

**Kata Kunci:** *Banjir, Fuzzy C-means, Clustering, Silhouette Coefficient*

## 1. Pendahuluan

Bencana alam merupakan peristiwa atau rangkaian peristiwa yang pengancam serta mengganggu kehidupan masyarakat yang mengakibatkan kerusakan lingkungan yang mengakibatkan adanya korban jiwa kerugian harta benda serta dampak psikologis yang mana penyebabnya berasal dari alam antara lain berupa banjir, tanah longsor, gempa bumi, tsunami, gunung meletus, kekeringan serta angin topan [1]. Bencana alam di Indonesia terjadi di setiap tahunnya mengakibatkan kerugian yang sangat besar. Bencana alam yang menjadi momok setiap tahunnya yaitu bencana banjir seluruh wilayah di Indonesia mengalami bencana ini termasuk di wilayah Jawa Timur yang rawan sekali terjadi bencana banjir.

Mitigasi merupakan rangkaian upaya untuk mengurangi resiko bencana, baik melalui pembangunan fisik maupun penyadaran dan peningkatan kemampuan menghadapi ancaman bencana [2]. Tindakan yang dapat dilakukan untuk mengurangi resiko banjir pada suatu wilayah salah satunya yaitu dengan tersedianya peta indeks rawan bencana.

Pemerintah Indonesia membentuk Badan Nasional Penanggulangan Bencana (BNPB) untuk merespon kejadian bencana yang terjadi. Dalam penyelesaian permasalahan bencana Badan Nasional Penanggulangan Bencana (BNPB) setiap wilayah satu dan lainnya diberikan pendekatan yang bersifat repetitive [3]. Pada faktanya, setiap wilayah memiliki karakteristik kejadian bencana yang tidak sama. Jika karakteristik disetiap kejadian wilayah diketahui, maka hal tersebut dapat mempermudah menyelesaikan dan menanggulangi masalah bencana yang cocok untuk wilayah wilayah tersebut. Dari penjelasan tersebut maka perlu diakukan pengelompokan wilayah kabupaten/kota di Indonesia berdasarkan kesamaan kejadian banjir yang telah terjadi sebelumnya, dengan harapan menghasilkan suatu pemetaan atau pengelompokan daerah rawan bencana banjir di wilayah kabupaten/kota di Jawa timur diharapkan nantinya dapat memberi masukan bagi BNPB atau pun pemerintah untuk menaggulangi bencana banjir dengan lebih efisien.

Beberapa penelitian yang dilakukan mengenai pengelompokan wilayah banjir seperti yang di lakukan oleh Pratikto R.O dengan penelitan yang berjudul Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir [4]. Dan juga penelitian yang dilakukan oleh Khomsiyah J.Dkk dengan penelitian yang berjudul Penerapan Algoritma K-Means Clustering untuk pengelompokan wilayah rawan banjir [5].

Sedangkan dalam penelitian ini metode yang digunakan yaitu Clustering *C-means* untuk pengelompokan bencana banjir pada kabupaten/kota di jawa timur. Metode Clustering *C-means* dipilih dikarenakan hasil nilai validasinya yang sangat baik dibandingkan hasil validasi metode *Clustering* lain [6]. Serta data yang dimiliki beserta variabelnya dapat dikelompokkan dalam *cluster cluster* sesuai

dengan kecenderungannya dan juga dengan metode ini dapat ditentukan jumlah *cluster* yang akan dibentuk kelebihan lainnya yaitu pusat clusternya berada ditempat yang lebih tepat dibanding metode lainnya [7]. Data yang digunakan pada penelitian ini merupakan data kejadian bencana banjir yang terjadi di Jawa Timur. Alasan melakukan pengelompokan ini yaitu untuk membantu pemerintah dalam pengambilan keputusan yang efisien dalam penanganan bencana alam banjir yang terjadi di Jawa timur untuk tiap wilayah yang memiliki karakteristik banjir yang berbeda dan agar pemerintah dapat memutuskan tindakan apa yang dapat dilakukan untuk mengantisipasi bencana banjir tersebut.

## 2. TINJAUAN PUSTAKA

### 2.1. Banjir

Banjir adalah peristiwa tergenangnya permukaan tanah yang biasanya diakibatkan oleh luapan air sungai ketika musim hujan atau dapat juga berasal dari kiriman dari wilayah lain yang permukaan tanahnya lebih tinggi. Indonesia sendiri, memiliki curah hujan yang tinggi berkisar antara 2000-3000 mm/tahun. Dan musim hujan di Indonesia biasanya terjadi pada bulan Oktober-Januari. Indonesia juga memiliki banyak sungai besar kurang lebih sebanyak 600 yang tersebar di berbagai wilayah. Namun, dikarenakan pengelolaan lingkungan yang kurang baik maka tak heran jika bencana banjir sering terjadi di Indonesia [8].

### 2.2. Fuzzy C-means

*Fuzzy C-means* (FCM), atau dikenal juga sebagai fuzzy Isodata merupakan salah satu metode clustering yang merupakan bagian dari metode *Hard K-Means*. FCM menggunakan model pengelompokan fuzzy sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1. Tingkat keberadaan data dalam suatu kelas atau *cluster* ditentukan oleh derajat keanggotaannya [9].

Berikut merupakan langkah – langkah dari algoritma *Fuzzy C-means*:

1. Input data yang akan di-*cluster*,  $X$ , berupa matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  data sampel ke- $i$  ( $i = 1, 2, \dots, n$ ), atribut ke- $j$  ( $j = 1, 2, \dots, m$ ).
2. Penentuan parameter yang akan digunakan, yaitu:
  - Jumlah *cluster* =  $c$
  - Pangkat =  $w$
  - Maksimum iterasi =  $MaxIter$
  - Error terkecil yang diharapkan =  $\epsilon$
  - Fungsi objektif awal =  $P_0 = 0$
  - Iterasi awal =  $t = 1$
3. Bangkitkan bilangan random  $\mu_{ik}$ ,  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; sebagai elemen-elemen matriks partisi awal  $U$ .

4. Hitung pusat *cluster* ke- $k$ ,  $V_{kj}$  dengan  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (2)$$

5. Hitung fungsi objektif pada iterasi ke- $t$ ,  $P_t$

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ([\sum_{j=1}^m (X_{ij} - V_{kj})^2] (\mu_{ik})^w) \quad (3)$$

6. Hitung perubahan matriks partisi

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (4)$$

dengan:  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$

7. Cek kondisi berhenti

- Jika:  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > MaxIter)$  maka berhenti;
- Jika tidak:  $t = t + 1$ , ulangi langkah ke-4.

### 2.3. Silhouette Coefficient

*Silhouette Coefficient* adalah metode yang digunakan untuk mengukur kualitas sebuah *cluster* yang menggabungkan nilai kohesi dan separasi. Nilai *silhouette coefficient* yang diperoleh terdapat pada rentang -1 hingga 1. Nilai *Silhouette Coefficient* yang mendekati 1 menunjukkan bahwa data tersebut semakin tepat berada di *cluster* yang dimaksud. Nilai *Silhouette Coefficient* negatif menunjukkan bahwa data tersebut tidak tepat berada dalam *cluster* yang dimaksud [10]. Kriteria subjektif pengukuran

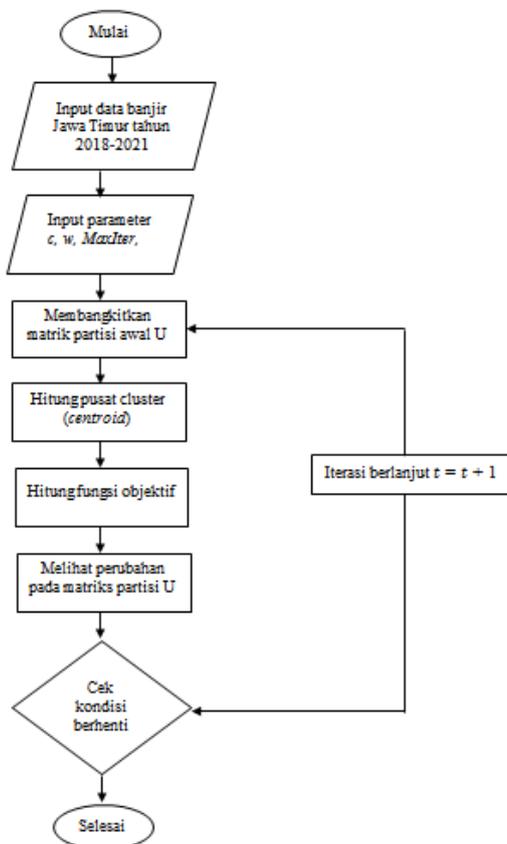
pengelompokan berdasarkan *Silhouette Coefficient* menurut Kauffman dan Roesseeuw (1990) dapat dilihat pada Tabel. 1.

Tabel 1. Kriteria pengukuran *Silhouette Coefficient*

Nilai	Kriteria
0,71-1,00	Struktur Kuat
0,51-0,70	Struktur Baik
0,26-0,50	Struktur Lemah
$\leq 0,25$	Struktur Buruk

### 3. METODOLOGI PENELITIAN

Adapun proses perhitungan algoritma *Fuzzy C-means* yang digunakan dalam penelitian ini ditunjukkan pada Gambar. 1 berikut:



Gambar 1. Alur penelitian

Data diperoleh dari web resmi BNPB (Badan Nasional Penanggulangan Bencana). Data yang digunakan meliputi jumlah kejadian bencana banjir, jumlah korban (korban meninggal, hilang, terluka, menderita, mengungsi kerusakan rumah), dan jumlah kerusakan fasilitas (rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas peribadatan, fasilitas umum, perkantoran, jembatan, pabrik, kios). Data yang diambil merupakan data bencana banjir di Jawa Timur dari bulan Januari 2018 sampai bulan Desember 2021 seperti pada Tabel. 2.

Tabel 2. Sample data yang akan digunakan

No.	Kabupaten/Kota	Jumlah Bencana Banjir	Korban	Kerusakan
1.	Banyuwangi	24	9721	517
2.	Batu	3	20	2
3.	Blitar	7	5000	21

4.	Bojonegoro	34	62141	65
5.	Bondowoso	2	1859	333
...	...	...	...	...
30.	Tulungagung	2	0	0

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Perhitungan Fuzzy C-means

Sebelum melakukan perhitungan dengan *Fuzzy C-means*, maka ditentukan terlebih dahulu parameter dari  $c$ ,  $w$ ,  $MaxIter$ , dan  $\epsilon$  seperti yang terdapat pada Tabel. 3.

Tabel 3. Parameter *Fuzzy C-means*

Parameter	Nilai
$C$	3
$W$	2
$MaxIter$	100
$\epsilon$	0,00001

Setelah dilakukan perhitungan untuk melakukan pengklusteran, diperoleh *centroid* akhir seperti pada Tabel.4 dan matriks partisi  $U$  seperti pada Tabel. 5.

Tabel 4. *Centroid*

Cluster	Centroid		
1	0.0003	2.7124	0.0001
2	0.0003	0.5877	0.0006
3	0.0001	0.0719	0.0005

Dari hasil matriks partisi  $U$  tersebut, akan dipilih nilai tertingginya untuk menentukan kelompok *cluster*.

Tabel 5. Hasil akhir *Fuzzy C-means*

No.	Kabupaten/Kota	U1	U2	U3	Cluster ke-
1.	Banyuwangi	0.0001	0.0027	0.9972	3
2.	Batu	0.0007	0.0147	0.9846	3
3.	Blitar	0.0001	0.0017	0.9983	3
4.	Bojonegoro	0.0003	0.9960	0.0037	2
5.	Bondowoso	0.0004	0.0087	0.9909	3
6.	Gresik	0.0015	0.9577	0.0407	2
7.	Jember	0.0011	0.9709	0.0280	2
8.	Jombang	0.0058	0.4984	0.4958	2
...	...	...	...	...	...
30.	Tulungagung	0.0007	0.0148	0.9846	3

Maka, hasil dari pengklusteran wilayah banjir dengan *Fuzzy C-means* pada Tabel. 5 diperoleh:

1. *cluster* 1 meliputi wilayah: Pasuruan.
2. *cluster* 2 meliputi wilayah: Bojonegoro, Gresik, Jember, Jombang, Lamongan.
3. *cluster* 3 meliputi wilayah: Banyuwangi, Batu, Blitar, Bondowoso, Kediri, Lumajang, Madiun, Magetan, Malang, Mojokerto, Nganjuk, Ngawi, Pacitan, Pamekasan, Ponorogo, Probolinggo, Sampang, Sidoarjo, Situbondo, Sumenep, Surabaya, Trenggalek, Tuban, Tulungagung.

4.2. Hasil

Berikut hasil dari analisis per *cluster* apabila dilihat dari variabel jumlah bencana banjir, korban, dan kerusakan seperti pada Tabel. 6.

Tabel 6. Nilai yang mendominasi

No.	Variabel	Cluster 1	Cluster 2	Cluster 3
1.	Jumlah Bencana Banjir	28	15-53	1-34
2.	Korban	274159	33017-126309	0-32535
3.	Kerusakan	9	9-96	0-517

Kemudian dari Tabel.6 dirata-rata dan diperoleh hasil seperti pada Tabel. 7.

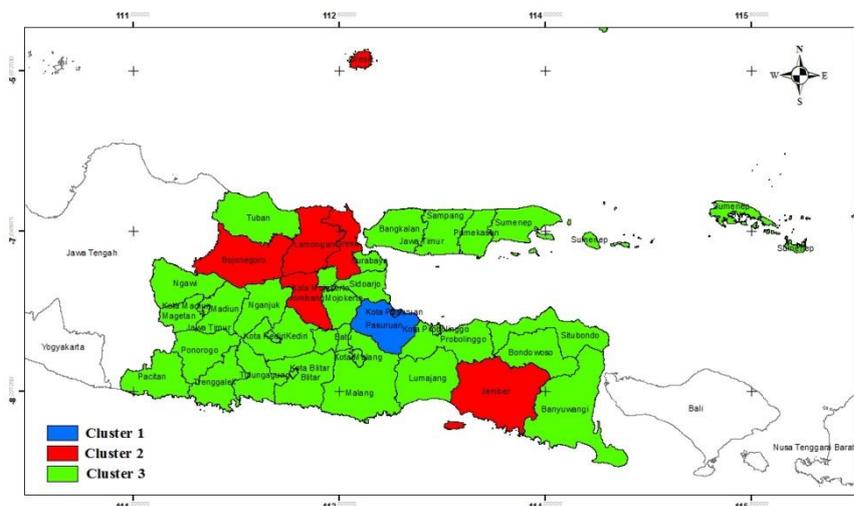
Tabel 7. Nilai rata-rata setiap *cluster*

No.	Variabel	Cluster 1	Cluster 2	Cluster 3
1.	Jumlah Bencana Banjir	28	31	10
2.	Korban	274159	64542	8267
3.	Kerusakan	9	54	51

Dapat dilihat pada Tabel. 7 bahwa pada *cluster 2*, nilai rata-rata pada variabel jumlah bencana banjir, korban, dan kerusakan memiliki nilai tertinggi yaitu sebanyak 31, 64542, dan 54. Lalu, untuk nilai rata-rata tertinggi kedua pada variabel jumlah bencana banjir terdapat pada *cluster 1* dan untuk nilai rata-rata tertinggi kedua pada variabel jumlah kerusakan terdapat pada *cluster 3*.

Berikut merupakan bentuk visualisasi dari hasil *clustering* berupa peta persebaran daerah rawan banjir seperti pada Gambar. 2.

PETA PENYEBARAN DAERAH RAWAN BANJIR DI PROVINSI JAWA TIMUR TAHUN 2018-2021



Gambar 2. Peta Persebaran Daerah Rawan Banjir Di Jawa Timur Tahun 2018-2021

4.3. Validasi Hasil Cluster

Dalam menentukan jumlah *cluster* terbaik dapat dilihat dari nilai *Silhouette Coefficient* dengan hasil yang diperoleh sebesar 0,84377. Apabila dilihat dari kriteria subjektif pengukuran pengelompokan menurut Kauffman dan Roesseeuw seperti pada Tabel. 1 maka dari hasil pengklusteran tersebut memiliki keakuratan yang kuat.

## 5. KESIMPULAN

Berdasarkan hasil dari algoritma *Fuzzy C-means* maka diperoleh: Wilayah pada *cluster* 1 meliputi: Pasuruan. Wilayah pada *cluster* 2 meliputi: Bojonegoro, Gresik, Jember, Jombang, Lamongan. Wilayah pada *cluster* 3 meliputi: Banyuwangi, Batu, Blitar, Bondowoso, Kediri, Lumajang, Madiun, Magetan, Malang, Mojokerto, Nganjuk, Ngawi, Pacitan, Pamekasan, Ponorogo, Probolinggo, Sampang, Sidoarjo, Situbondo, Sumenep, Surabaya, Trenggalek, Tuban, Tulungagung. Pada *cluster* 2, nilai rata-rata pada variabel jumlah bencana banjir, korban, dan kerusakan memiliki nilai tertinggi yaitu sebanyak 31, 64542, dan 54. Lalu, untuk nilai rata-rata tertinggi kedua pada variabel jumlah bencana banjir terdapat pada *cluster* 1 dan untuk nilai rata-rata tertinggi kedua pada variabel jumlah kerusakan terdapat pada *cluster* 3. Hasil dari clustering *Fuzzy C-means* dengan nilai  $c = 3$ ,  $w = 2$  merupakan hasil *cluster* terbaik karena hasil dari validasi dengan menggunakan *Silhouette Coefficient* sebesar 0,84377.

## DAFTAR PUSTAKA

- [1] BNPB, "Definisi Bencana," *bnpb.co.id*, 2007. <https://www.bnpb.go.id/definisi-bencana>.
- [2] D. F. Niode, Y. D. Y. Rindengan, and S. D. . Karouw, "Geographical Information System (GIS) untuk Mitigasi Bencana Alam Banjir di Kota Manado," *J. Tek. Elektro dan Komput.*, vol. 5, no. 2, pp. 14–20, 2016.
- [3] A. Ramadhan, Mustakim, and R. Handinata, "Implementasi Algoritma Fuzzy C Means Dan Moora Untuk Pengelompokan Dan Penentuan Wilayah Penanggulangan Bencana Banjir," no. November, p. Pekanbaru, 2019.
- [4] R. O. Pratiko and N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," *JOINTECS (Journal Inf. Technol. Comput. Sci.*, vol. 6, no. 1, p. 13, 2021, doi: 10.31328/jointecs.v6i1.1473.
- [5] P. Algoritma and K. C. Untuk, "Jahrotun Khomsiyah, Alif Ramdhani, Ade Feby Damayanti, Dede Rohman PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKAN WILAYAH RAWAN BANJIR Jahrotun Khomsiyah 1 , Alif Ramdhani 2 , Ade Feby Damayanti 3 , Dede Rohman 4," no. 03, pp. 249–253, 2021.
- [6] A. Ramadhan, Z. Efendi, and Mustakim, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 9, pp. 219–226, 2017.
- [7] F. Agustini, "Implementasi Algoritma Fuzzy C-Means Studi Kasus," *J. Ilmu Pengetah. Dan Teknol. Komput.*, vol. 3, no. 1, pp. 127–132, 2017.
- [8] P. Kegeografian and D. I. K. Semarang, "Kesiap Siagaan Masyarakat Dalam Penanggulangan Banjir Di Kota Semarang," *J. Geogr. Media Inf. Pengemb. dan Profesi Kegeografian*, vol. 12, no. 1, pp. 102–114, 2018, doi: 10.15294/jg.v12i1.8019.
- [9] D. L. Rahakbauw, V. Y. I. Ilwaru, and M. H. Hahury, "Implementasi Fuzzy C-Means Clustering Dalam Implementation Of Fuzzy C-Means Clustering In," *J. Ilmu Mat. dan Terap.*, vol. 11, pp. 1–12, 2017.
- [10] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.