

Analisis *Clustering* Daerah Rawan Banjir Menggunakan *Fuzzy C-Means* (Studi Kasus : Bakorwil 2 Jawa Timur)

Monike Febriyani Faris¹, Putroue Keumala Intan², Widodo³

¹Universitas Islam Negeri Sunan Ampel Surabaya, monikefebriyani@gmail.com

²Universitas Islam Negeri Sunan Ampel Surabaya, putroue@uinsby.ac.id

³BPBD Jawa Timur, mail@bpbj.jatimprov.go.id

Abstrak: Banjir dapat membahayakan kehidupan manusia, seperti dapat merusak rumah warga, merusak fasilitas umum atau infrastruktur dan lahan pertanian milik warga. Kajian daerah rawan bencana banjir dapat menjadi acuan untuk menentukan kemungkinan besarnya kerugian serta dapat menjadi serangkaian upaya pemerintah untuk mengurangi risiko bencana menjadi lebih efektif. Salah satu daerah di Indonesia yang rawan terjadi banjir adalah Bakorwil 2 Jawa Timur, terlebih ketika musim penghujan. Untuk memetakan atau mengelompokkan daerah rawan banjir dalam penelitian ini menggunakan *Fuzzy C-Means*. Hasil pengklasteran menunjukkan bahwa 94 kecamatan di Bakorwil 2 Jawa Timur paling baik diklaster menjadi 5 *cluster*, terdapat 6 kecamatan yang termasuk kedalam daerah rawan dengan risiko banjir sangat tinggi dan sebanyak 63 kecamatan termasuk kedalam daerah rawan banjir dengan risiko sangat rendah.

Kata kunci: *Banjir, Daerah Rawan, Bakorwil 2 Jawa Timur, Fuzzy C-Means*

Abstract: *Floods can endanger human life, such as damaging people's houses, damaging public facilities or infrastructure and residents' agricultural land. The study of flood-prone areas can be a reference to determine the possible magnitude of losses and can be a series of government efforts to reduce disaster risk to be more effective. One area in Indonesia that is prone to flooding is Bakorwil 2 East Java, especially during the rainy season. To map or classify flood-prone areas in this study using Fuzzy C-Means. Clustering results show that 94 sub-districts in Bakorwil 2 East Java are best clustered into 5 clusters, there are 6 sub-districts which are classified as vulnerable areas with very high risk of flooding and as many as 63 sub-districts are included in flood-prone areas with very low risk.*

Keywords: *Flood, Prone Areas, Bakorwil 2 East Java, Fuzzy C-Means*

1. Pendahuluan

Negara Indonesia merupakan salah satu negara yang rawan bencana. Kondisi iklim dan kondisi topografi permukaan tanah yang dimiliki Indonesia membuat kondisi tanah Indonesia subur. Tetapi, kondisi tersebut juga dapat menimbulkan beberapa bencana seperti banjir, tanah longsor, kekeringan dan kebakaran hutan [1].

Wilayah Jawa Timur terbagi dalam 5 Bakorwil (Badan Koordinasi Wilayah), salah satunya adalah Bakorwil 2 yang terdiri dari Kabupaten Mojokerto, Kota Mojokerto, Kabupaten Lamongan, Kabupaten Bojonegoro, Kabupaten Tuban, Kabupaten Kediri, Kota Kediri, Kabupaten Jombang. Ada beberapa ancaman bencana di Bakorwil 2 Jawa Timur salah satunya adalah bencana banjir. Terlebih ketika musim penghujan, banjir merupakan salah satu bencana yang identik dengan musim hujan. Banjir adalah bencana akibat curah hujan tinggi yang menyebabkan aliran sungai meluap sehingga menggenangi suatu wilayah yang lebih rendah di sekitarnya [2]. Banjir dapat membahayakan kehidupan manusia jika banjir memasuki wilayah pemukiman warga. Banjir dapat merusak rumah warga, merusak fasilitas umum atau infrastruktur dan lahan pertanian milik warga.

Tingginya risiko bencana banjir mengharuskan pemerintah untuk memberikan perlindungan dari ancaman bencana. Kajian daerah rawan bencana berdasarkan dampak bencana yang sudah terjadi dapat menjadi salah satu acuan untuk menentukan kemungkinan besarnya kerugian dan kerusakan akibat ancaman bencana. Sehingga, pemerintah dapat menyelenggarakan mitigasi bencana yang merupakan serangkaian upaya untuk mengurangi risiko bencana menjadi lebih efektif.

Dalam mengelompokkan atau memetakan daerah rawan bencana dapat menggunakan *fuzzy clustering*. *Fuzzy clustering* merupakan teknik pengelompokan secara optimal berdasarkan logika *fuzzy*. *Fuzzy clustering* mempunyai beberapa metode, salah satunya adalah *Fuzzy C-Means*. *Fuzzy C-Means* (FCM) adalah metode pengklasteran yang terawasi, karena perlu diketahui jumlah *cluster* yang akan dibentuk [3]

Terdapat beberapa penelitian terdahulu yang juga menggunakan metode *fuzzy c-means*, dimana metode tersebut dapat memetakan atau mengelompokkan secara optimal. Seperti pada jurnal penelitian “*Surabaya Tourism Destination Recommendation Using Fuzzy C-Means Algorithm*”, hasil aplikasi penelitian tersebut dapat memberikan rekomendasi tentang destinasi turis yang akurat sesuai yang dibutuhkan para turis [4]. Selain itu, terdapat penelitian lain yang berjudul “*Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C-Means Clustering*”, penelitian tersebut menunjukkan bahwa hasil *cluster* dari *fuzzy c-means* lebih baik dari *fuzzy k-means* [5]. Terdapat penelitian terdahulu yang berhubungan dengan pengklasteran daerah risiko banjir yaitu jurnal penelitian yang berjudul “*Risk Analysis of Flood Disaster Based on Fuzzy Clustering Method*”, hasil penelitian tersebut mengelompokkan 30 provinsi di Cina menjadi 5 kelompok berdasarkan risiko banjir [6].

2. Kajian Pustaka

2.1 Banjir

Bencana yang sering terjadi ketika musim hujan di Indonesia adalah banjir. Banjir adalah bencana akibat curah hujan tinggi yang menyebabkan aliran sungai meluap sehingga menggenangi suatu wilayah yang lebih rendah di sekitarnya dan dalam jangka waktu tertentu [2].

Banjir dapat membahayakan kehidupan manusia jika banjir memasuki wilayah pemukiman warga. Banjir dapat menimbulkan kerusakan, seperti rusaknya rumah

penduduk, rusaknya fasilitas umum atau infrastruktur, rusaknya lahan atau persawahan penduduk. Jenis banjir dapat dibedakan menjadi tiga macam. Ketiga jenis tersebut adalah

- a. **Banjir Bandang**
Banjir bandang adalah banjir besar yang berlangsung sesaat dan secara tiba-tiba. Banjir bandang diakibatkan oleh curah hujan yang sangat tinggi dengan jangka waktu pendek sehingga menyebabkan debit sungai naik secara cepat. Selain curah hujan tinggi, banjir bandang dapat disebabkan oleh kondisi geologi, morfologi dan tutupan lahan.
- b. **Banjir Sungai**
Banjir sungai adalah banjir yang disebabkan oleh hujan yang terjadi di daerah sekitar sungai secara luas dan dengan jangka waktu lama, sehingga air sungai meluap dan menggenangi daerah disekitar sungai. Banjir sungai umumnya terjadi secara perlahan dan berlangsung lama sampai berhari-hari bahkan berminggu-minggu.
- c. **Banjir Rob**
Banjir rob adalah banjir yang terjadi akibat air laut yang pasang atau gelombang badai sehingga air laut menggenangi daratan yang lebih rendah di sekitar pantai.

2.2 Rawan Bencana

Kerawanan daerah terhadap bencana tertentu perlu penilaian berupa indeks kerawanan bencana dimana pihak BNPB telah melakukan penilaian indeks kerawanan bencana per provinsi di Indonesia tahun 2009, 2011 dan 2013. Penilaian kerawanan bencana merupakan penilaian tingkat kerusakan atau kerugian yang diakibatkan oleh bencana tertentu. Parameter yang diperlukan dalam penilaian kerawanan daerah terhadap bencana antara lain kerusakan rumah, kerusakan fasilitas umum atau infrastruktur, korban meninggal, korban luka-luka serta kepadatan penduduk [7].

Daerah disebut daerah dengan kerawanan tinggi terhadap bencana tertentu ketika frekuensi kejadian bencana tersebut banyak serta kerusakan dan kerugian yang terdampak besar. Sedangkan, daerah rawan rendah jika frekuensi serta kerusakan dan kerugian dampak dari bencana tersebut sedikit.

2.3 Fuzzy C-Means

Fuzzy C-Means (FCM) merupakan salah satu teknik pengklasteran data dimana keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan [3]. FCM diperkenalkan oleh Jim Bezdek pada tahun 1981. Konsep dasar *Fuzzy C-Means*, pertama kali yaitu menentukan pusat *cluster* yang akan menandai lokasi rata-rata untuk masing-masing *cluster* [8]. Pada kondisi awal, pusat *cluster* masih belum akurat sehingga dibutuhkan perbaikan pusat *cluster* dan derajat keanggotaan masing-masing titik data secara berulang hingga berada pada titik yang tepat. *Output* dari FCM berupa deretan pusat *cluster* dan derajat keanggotaan tiap titik data. *Output* tersebut dapat digunakan untuk membangun suatu *fuzzy inference system*.

Algoritma *Fuzzy C-Means* sebagai berikut [3],

1. Memasukkan data yang akan di hitung dalam bentuk matriks $n \times m$.
 n menunjukkan jumlah sampel data, dan m menunjukkan atribut dari setiap data.
 x_{ij} = data sampel ke i ($i = 1, 2, \dots, n$), atribut ke ke j ($j = 1, 2, \dots, m$)
2. Menentukan jumlah *cluster* (k), pembobot (w) dengan $w > 1$, iterasi maksimal, fungsi objektif awal ($p_0 = 0$), error terkecil yang diharapkan (ξ) serta iterasi awal (t).
3. Membangkitkan matriks random awal U , berupa nilai acak atau random μ_{ik} , dengan $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$. μ_{ik} adalah derajat keanggotaan yang

menunjukkan seberapa besar kemungkinan data menjadi anggota ke dalam suatu *cluster*, dimana derajat keanggotaannya terletak pada interval 0 sampai 1, dan jika dijumlahlah tiap baris akan berjumlah 1.

$$Q_i = \sum_{k=1}^c \mu_{ik} \quad (1)$$

4. Menghitung pusat *cluster* ke-k (V_{kj})

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

5. Menghitung fungsi objektif pada iterasi ke-t (P_t)

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

6. Menghitung perubahan matriks partisi

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

7. Kemudian mengecek kondisi

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

Sebelum data masuk dalam perhitungan FCM, data dinormalisasi terlebih dahulu dikarenakan jarak antara titik data yang digunakan pada penelitian cukup besar. Sehingga data tersebut harus dinormalisasi supaya tidak ada variabel yang mendominasi dalam perhitungan jarak antar data. Pada penelitian ini menggunakan normalisasi min-max.

$$x_{ij} = \frac{x_{min_{ij}} - x_{min_j}}{x_{max_j} - x_{min_j}} \quad (5)$$

$i = 1, 2, 3, \dots, n$; $j = 1, 2, 3, \dots, m$

2.4 Silhouette Index (SI)

Silhouette Index (SI) digunakan untuk memvalidasi sebuah data, *cluster* tunggal (satu *cluster* dari sejumlah *cluster*), atau bahkan seluruh *cluster*, dengan kata lain metode ini digunakan untuk menguji kualitas tiap *cluster* [9]. Perhitungan nilai *Silhouette Index* dari sebuah data ke- i , ada 2 komponen yaitu a_i dan b_i . a_i merupakan rata-rata jarak ke- i terhadap semua data lainnya dalam satu *cluster*. Sedangkan b_i diperoleh dengan menghitung rata-rata jarak data ke- i terhadap semua data dari *cluster* lainnya, kemudian mengambil yang terkecil.

Dalam menghitung nilai *SI* membutuhkan nilai jarak antar data, rumus yang digunakan dalam menghitung jarak antar data adalah *Squared Euclidian Distance*,

$$d(p, q) = \sum_{i=1}^n (q_i - p_i)^2 \quad (6)$$

1. Untuk menghitung a_i adalah sebagai berikut :

$$a_i^j = \frac{1}{m_j - 1} \sum_{r=1}^{m_j} d(x_i^j, x_r^j) \quad (7)$$

$i = 1, 2, \dots, m_j$

$d(x_i^j)(x_r^j)$ = jarak data ke- i dengan data ke- r dalam satu *cluster*

m_j = jumlah data dalam *cluster* ke- j

2. Untuk menghitung b_i adalah sebagai berikut :

$$b_i^j = \min \left\{ \frac{1}{m_n} \sum_{r=1}^{m_n} d(x_i^j, x_r^n) \right\} \quad (8)$$

$$i = 1, 2, \dots, m_n$$

3. Untuk mendapatkan *Silhouette Index* (SI) data ke-*i* menggunakan persamaan berikut :

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}} \quad (9)$$

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \quad (10)$$

dimana :

$$SI_i^j = SI \text{ data ke-}i$$

$$SI_j = SI \text{ untuk tiap cluster}$$

Rentang nilai *Silhouette Index* berada dalam rentang [-1, +1]. Nilai *Silhouette Index* mendekati 1 menunjukkan bahwa data tersebut semakin tepat berada pada *cluster* tersebut. Sedangkan nilai *Silhouette Index* negatif menandakan bahwa data tersebut tidak tepat berada pada *cluster* tersebut karena lebih dekat dengan *cluster* yang lain. *Silhouette Index* bernilai 0 atau mendekati 0 menandakan data tersebut berada pada posisi perbatasan antara dua *cluster*.

Menurut Kaufman dan Rousseeuw, berikut merupakan ukuran nilai *silhouette coefficient* (SC) [10].

- a. $0,7 < SC \leq 1$, berarti *strong structure*
- b. $0,5 < SC \leq 0,7$, berarti *medium structure*
- c. $0,25 < SC \leq 0,5$, berarti *weak structure*
- d. $SC \leq 0,25$, berarti *no structure*

3. Metode Penelitian

3.1 Jenis Penelitian

Jenis penelitian ini adalah penelitian kuantitatif, penelitian dimana datanya berupa angka-angka dan dianalisis berdasarkan rumus statistik, dimana datanya didapat dari suatu pengamatan dengan suatu alat bantu [11].

3.2 Pengumpulan Data

Pengumpulan data dengan 2 cara yaitu mencari bukti data dari sumber yang terkait dengan penelitian dan wawancara dengan pihak terkait agar bisa melengkapi dan memperjelas data.

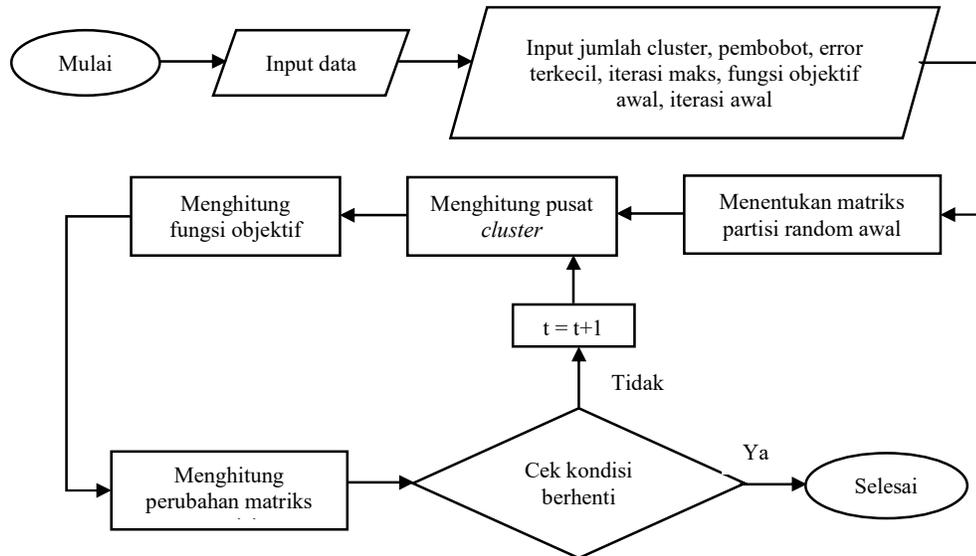
Data yang digunakan adalah 94 data kecamatan yang terdampak banjir di Bakorwil 2 Jawa Timur dari tahun 2014-2018 yang diperoleh dari SmartPB BPBD (Badan Penanggulangan Bencana Daerah) Jawa Timur. Data tersebut berisi frekuensi banjir, jumlah rumah rusak, jumlah rumah tergenang, luas lahan pertanian tergenang, fasilitas umum tergenang dan jumlah korban bencana banjir.

Tabel 1. Tabel Sampel Data Banjir Bakorwil 2 2014-2018

Kode Kec	Nama Kecamatan	Frekuensi Banjir	Rumah Tergenang	Rumah Rusak	Sawah Tergenang (Ha)	Fasilitas Umum Tergenang	Korban
350602	Mojo	2	120	0	0	0	0
351605	Ngoro	10	570	0	32	0	0
351709	Jombang	6	50	0	85	0	0
352317	Plumpang	6	2567	0	526	61	1
357102	Kediri	2	50	0	0	0	0

3.3 Pengolahan Data

Data tersebut akan diklaster menjadi 3 kategori daerah rawan banjir dengan risiko rendah, sedang, dan tinggi. Proses pengolahan data terdapat beberapa tahap. Pertama, data yang sudah ada dinormalisasi terlebih dahulu agar tidak ada variabel yang mendominasi. Kemudian data diproses menggunakan metode *fuzzy c-means* dengan *flowchart* pada gambar 1. Setelah data diproses dengan *fuzzy c-means*, hasil *cluster* dari data tersebut akan dianalisis dengan *Silhouette index* untuk melihat kualitas *cluster*.



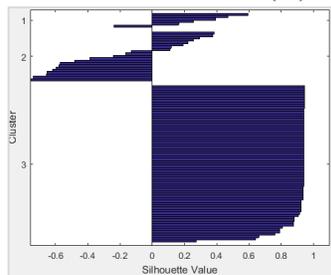
Gambar 1. *Flowchart Fuzzy C-Means*

4. Hasil dan Pembahasan

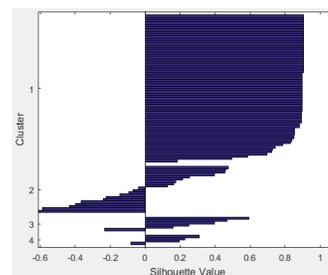
Pada penelitian ini terdapat 94 data kecamatan di Bakorwil 2 Jawa Timur yang terdampak bencana banjir. Data tersebut akan dikelompokkan dalam 3, 4, 5 dan 6 *cluster* yaitu daerah rawan banjir dengan risiko rendah, sedang dan tinggi. Sebelum data diolah menggunakan FCM, data terlebih dahulu dinormalisasi menggunakan rumus normalisasi min-max pada persamaan (2.5).

Setelah 94 data dinormalisasi, data tersebut diolah menggunakan *fuzzy c-means*. Sebelum memasuki perhitungan *fuzzy c-means*, langkah pertama yaitu menentukan jumlah *cluster*, pembobot, maksimum iterasi dan error terkecil. Pembobot (w) = 2, maksimum iterasi = 100, error terkecil (ξ) = 0,000001, fungsi objektif awal (P_0) = 0.

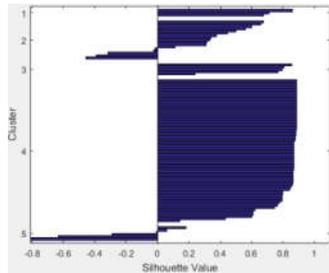
Hasil tiap *cluster* akan diuji validalitas untuk menilai seberapa kuat *cluster* menggunakan *Silhouette Index (SI)*.



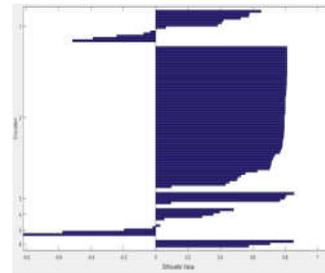
Gambar 2. *Silhouette Index 3 Cluster*



Gambar 3. *Silhouette Index 4 Cluster*



Gambar 4. *Sillhouette Index 5 Cluster*



Gambar 5. *Sillhouette Index 6 Cluster*

Tabel 2. Hasil Perbandingan *SI Global*

<i>SI Cluster 3</i>	0,319718
<i>SI Cluster 4</i>	0,316701
<i>SI Cluster 5</i>	0,451014
<i>SI Cluster 6</i>	0,380536

Berdasarkan hasil perhitungan, jika data dibagi menjadi 3 *cluster* terdapat 14 data yang *overlapping*, 1 data dari *cluster* 1 dan 13 data dari *cluster* 2. Untuk data yang dibagi menjadi 4 *cluster* sebanyak 13 data yang *overlapping* nilai *SI global* nya sebesar 0,316701. Untuk data yang dikelompokkan menjadi 5 *cluster* terdapat beberapa data yang *overlapping* antara lain pada data 1, 3, 5 dan 6 dalam *cluster* 5 dan data 7, 8, 12, 14, 17 pada *cluster* 2. Sehingga, terdapat 9 data yang nilai *SI* nya bernilai negatif dengan presentasi sekitar 90% data bernilai positif. Sedangkan, jika data dikelompokkan menjadi 6 *cluster*, jumlah data yang *overlapping* sebanyak 9 data dengan nilai *SI global* sebesar 0,3805356. Jika dibandingkan ketika 94 data dibagi menjadi 3, 4, 5 dan 6 *cluster*, jumlah data *overlapping* yang paling sedikit adalah ketika jumlah *cluster* 5 yaitu sebanyak 9 data. Selain itu, nilai *SI global* ketika jumlah *cluster* 5 paling mendekati 1 yaitu sebesar 0,451014. Sehingga, 94 data kecamatan daerah rawan banjir paling baik dibagi menjadi 5.

Selanjutnya menghitung *mean* dari keenam variabel untuk tiap *cluster* yang terbentuk. Nilai *mean* yang lebih tinggi menandakan kerugian dan dampak banjir yang lebih besar. Berdasarkan hasil perhitungan *mean* menunjukkan bahwa:

- a. *Cluster* 1 terdapat 3 kecamatan yang merupakan daerah rawan banjir risiko tinggi dengan frekuensi banjir terbanyak dari kecamatan lain
- b. *Cluster* 2 merupakan daerah rawan dengan risiko banjir level sedang dimana terdapat 17 Kecamatan termasuk dalam *cluster* 2
- c. *Cluster* 3 terdapat 5 Kecamatan yang merupakan daerah rawan dengan risiko banjir level rendah
- d. *Cluster* 4 merupakan daerah rawan dengan risiko banjir level sangat rendah sebanyak 63 Kecamatan
- e. *Cluster* 5 terdapat 6 kecamatan yang merupakan daerah rawan dengan risiko banjir level sangat tinggi.

5. Kesimpulan

Berdasarkan pengolahan dan analisis data pada bab sebelumnya, maka kesimpulan yang didapat adalah

- a. Pengklasteran 94 data kecamatan di Bakorwil 2 Jawa Timur, dibagi menjadi 5 *cluster* berdasarkan kerugian dan dampak banjir. *Cluster* 1 (Daerah rawan banjir dengan risiko tinggi) terdapat 3 kecamatan, untuk *Cluster* 2 (Daerah rawan dengan risiko banjir level sedang) terdapat 17 Kecamatan, *Cluster* 3 (Daerah rawan dengan

risiko banjir level rendah) terdapat 5 Kecamatan, *Cluster 4* (Daerah rawan dengan risiko banjir level sangat rendah) terdapat 63 Kecamatan, dan *Cluster 5* (Daerah rawan dengan risiko banjir level sangat tinggi) terdapat 6 Kecamatan.

- b. Hasil *Silhouette index* jika data dibagi menjadi 5 *cluster* yaitu nilai *SI* pada *cluster 1* adalah 0,747466, *cluster 2* sebesar 0,236415, *cluster 3* sebesar 0,693167, untuk *cluster 4* sebesar 0,82862, sedangkan untuk *SI cluster 5* adalah -0,2506, dan jumlah data yang *overlapping* sebanyak 9 data.

Daftar Pustaka

- [1] BNPB, “BNPB,” 24 Januari 2019. [Online]. Available: <http://www.bnpb.go.id/home/potensi>.
- [2] Tim Penulis BNPB, Buku Saku Tanggap Tangkas Tangguh Menghadapi Bencana, Yogyakarta, 2017.
- [3] S. Kusumadewi dan H. Purnomo, Aplikasi Logika Fuzzy, 2 penyunt., Yogyakarta: Graha Ilmu, 2010.
- [4] R. Sutjiadi, E. M. Trianto dan A. G. Budihardjo, “Surabaya Tourism Destination Recommendation Using Fuzzy C-Means Algorithm,” *Journal of Telecommunication, Electronic and Computer Engineering*, vol. 10, pp. 177-181, 2018.
- [5] S. C. Sripada dan M. S. Rao, “Comparison of Purity and Entropy of K-Means Clustering and Fuzzy C-Means Clustering,” *Indian Journal of Computer Science and Engineering*, vol. 2, pp. 343-346, 2011.
- [6] J. Chen, S. Zhao dan H. Wang, “Risk Analysis of Flood Disaster Based on Fuzzy Clustering Method,” *Energy Procedia*, pp. 1915-1919, 2011.
- [7] L. Kurniawan, S. Triutomo, R. Yunus, M. R. Amri dan A. A. Hantyanto, Indeks Rawan Bencana, Sentul: Direktorat Pengurangan Risiko Bencana Deputi Bidang Pencegahan dan Kesiapsiagaan, 2011.
- [8] Sujarwata, Sistem Fuzzy dan Aplikasinya, Semarang: UNS, 2014.
- [9] V. T. P. Swindiarto, R. Sarno dan D. C. R. Novitasari, “Integration of Fuzzy C-Means Clustering and TOPSIS (FCM-TOPSIS) with Silhouette Analysis for Multi Criteria Parameter Data,” *International Semantic*, pp. 463-468, 2018.
- [10] L. Kaufman dan P. J. Rousseeuw, Finding Groups in Data, New York: John Wiley & Sons, 1990.
- [11] N. Alfianika, Buku Ajar Metode Penelitian Pengajaran Bahasa Indonesia, Yogyakarta: Deepublish, 2016.