

Received: 14 October 2022  
Revised: 21 December 2022  
Accepted: 29 December 2022  
Published: 31 December 2022

## Penggerombolan Desa di Jawa Barat Berdasarkan Daerah Rawan Bencana

Defri Ramadhan Ismana<sup>1, a)</sup>, Seta Baehera<sup>1, b)</sup>, Anwar Fitrianto<sup>1, c)</sup>,  
Bagus Sartono<sup>1, d)</sup>, Sachnaz Desta Oktarina<sup>1, e)</sup>

<sup>1</sup>*Program Studi Statistika dan Sains Data, IPB University*

E-mail: <sup>a)</sup>dr.ismana@apps.ipb.ac.id, <sup>b)</sup>setabaehera@apps.ipb.ac.id, <sup>c)</sup>anwarstat@gmail.com,  
<sup>d)</sup>bagusco@apps.ipb.ac.id, <sup>e)</sup>sachnazdes@apps.ipb.ac.id

### Abstract

Indonesia is one of the countries that has a large potential for natural disasters. Indonesia's position at the confluence of 4 continental plates makes the potential for earthquakes even greater. The tropical climate with 2 seasons makes changes in weather, temperature and wind direction quite extreme. These climatic conditions combined with the relatively diverse surface and rock topography conditions, these conditions can cause several bad consequences for the community such as hydrometeorological disasters such as floods, landslides, forest fires, and droughts. Particularly in West Java province, natural disasters that have occurred include: landslides, droughts, cyclones/typhoons, tidal waves, fires, volcanic eruptions, tsunamis, and other disasters. The purpose of this study was to cluster villages in the West Java region based on the level of disaster-prone in 2018. The research was carried out using K-Prototypes clustering and testing evaluation using the silhouette coefficient. The results showed that the optimal number of clusters in this study was nine clusters. These clusters can be distinguished based on the disaster category and the characteristics of the area.

**Keywords:** clustering, K-Prototypes, silhouette coefficient

### Abstrak

Indonesia merupakan salah satu negara yang memiliki potensi bencana alam yang cukup besar. Posisi Indonesia yang berada pertemuan 4 lempeng benua membuat potensi bencana gempa bumi kian besar. Iklim tropis dengan 2 musim membuat perubahan cuaca, suhu dan arah angin yang cukup ekstrim. Kondisi iklim seperti ini digabungkan dengan kondisi topografi permukaan dan batuan yang relatif beragam, kondisi ini dapat menimbulkan beberapa akibat buruk bagi masyarakat seperti terjadinya bencana hidrometeorologi seperti halnya banjir, tanah longsor, kebakaran hutan dan kekeringan. Khususnya provinsi Jawa Barat, bencana alam yang pernah terjadi antara lain: tanah longsor, kekeringan, angin puyuh/puting beliung/topan, gelombang pasang, kebakaran, gunung meletus, tsunami dan bencana lainnya. Tujuan penelitian ini adalah menggerombolan desa-desa yang berada pada wilayah Jawa Barat berdasarkan tingkat rawan bencana pada tahun 2018. Metode penelitian dilakukan menggunakan penggerombolan *K-Prototypes* dan evaluasi pengujian menggunakan koefisien *silhouette*. Hasil penelitian ini menunjukkan bahwa jumlah gerombol yang optimal adalah 9. Gerombol-gerombol tersebut dapat dibedakan berdasarkan kategori bencana dan karakteristik wilayahnya.

**Kata-kata kunci:** penggerombolan, *K-Prototypes*, koefisien *silhouette*

## PENDAHULUAN

Secara geologis Indonesia merupakan Negara kepulauan yang terletak pada pertemuan empat lempeng utama, yaitu lempeng Eurasia, Indo Australia, Filipina dan Pasifik. Pada bagian selatan dan timur Indonesia terdapat sabuk vulkanik yang memanjang dari Pulau Sumatera, Jawa, Nusa Tenggara hingga Sulawesi Sulawesi. Kondisi ini menjadikan Indonesia rawan bencana gempa bumi, tsunami, dan letusan gunung api. Secara geografis wilayah Indonesia terletak di daerah iklim tropis dengan dua musim yaitu panas dan hujan dengan ciri-ciri adanya perubahan cuaca, suhu dan arah angin yang cukup ekstrim. Kondisi iklim seperti ini digabungkan dengan kondisi topografi permukaan dan batuan yang relatif beragam, baik secara fisik maupun kimiawi, menghasilkan kondisi tanah yang subur. Sebaliknya, kondisi ini dapat menimbulkan beberapa akibat buruk bagi masyarakat seperti terjadinya bencana hidrometeorologi seperti halnya banjir, tanah longsor, kebakaran hutan dan kekeringan (Adi, et al., 2021).

Berdasarkan data BNPB, pada tahun 2019 hingga tahun 2021 terjadi 3,006 kejadian bencana alam di provinsi Jawa Barat yang merupakan provinsi dengan jumlah kejadian bencana alam tertinggi di Indonesia. Adapun beberapa bencana alam yang terjadi di Jawa Barat yaitu tanah longsor, kekeringan, angin puyuh/puting beliung/topan, gelombang pasang, kebakaran, gunung meletus, tsunami dan bencana lainnya. Berdasarkan data dari Open Data Jabar, portal resmi data terbuka milik Pemda Provinsi Jawa Barat, dari berbagai macam kategori bencana alam yang ada di Jawa Barat, tanah longor merupakan bencana alam yang sering terjadi di Jawa Barat. Selama tahun 2020 sebanyak 3.312 kasus tanah longsor terjadi di Jawa Barat. Bencana alam lainnya yang sering terjadi adalah gempa bumi, banjir, kebakaran, dan lainnya. Garut merupakan wilayah yang cukup sering terjadi bencana alam. Bencana alam yang terjadi di Garut mencapai 1.491 kejadian dengan bencana alam yang paling sering terjadi yaitu Tanah Longsor sebanyak 505 kejadian.

Beberapa penelitian mengenai bencana alam di wilayah Indonesia telah dilakukan sebelumnya. Penelitian yang dilakukan oleh Linda Tondobala (2012) mengenai pendekatan untuk menentukan kawasan rawan bencana di pulau Sulawesi. Faizana et al (2015) mengenai pemetaan risiko bencana tanah longsor di Kota Semarang. Triyanto et al (2014) tentang analisis klasifikasi bencana banjir berdasarkan curah hujan menggunakan algoritma *Naïve Bayes*. Kemudian penerapan algoritma *clustering* dalam mengelompokkan banyaknya desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam menurut provinsi dengan *K-Means* oleh Sardewo et al (2018).

Tujuan penelitian ini adalah menggerombolkan desa-desa yang berada pada wilayah Jawa Barat berdasarkan tingkat rawan bencana pada tahun 2018. Dengan adanya penggerombolan ini dapat kita lihat sebaran wilayah di Jawa Barat yang memiliki tingkat rawan bencana yang cukup tinggi. Pada penelitian ini data yang digunakan adalah data Potensi Desa (Podes) tahun 2018 yang dipublikasikan oleh Badan Pusat Statistik (BPS). Cakupan penelitian ini kami batasi hanya pada provinsi Jawa Barat, karena provinsi Jawa Barat merupakan provinsi terpadat penduduknya di Indonesia. Informasi yang digunakan pada penelitian ini yaitu data kejadian bencana alam dan jumlah korban. Penggerombolan terhadap desa-desa di Jawa Barat menggunakan metode *K-Prototypes*. *K-Prototypes* merupakan suatu algoritma analisis *cluster* yang dapat digunakan untuk mengelompokkan objek berdasarkan data campuran yaitu data numerik dan kategorik. Selanjutnya metode koefisien *silhouette* digunakan untuk mengevaluasi grombol yang telah dibuat.

## METODOLOGI

### Data

Data yang digunakan pada penelitian ini adalah data Potensi Desa (Podes) tahun 2018 yang dipublikasikan oleh Badan Pusat Statistik (BPS). Unit yang penelitian yang digunakan sebanyak 5,957 desa di provinsi Jawa Barat. Informasi yang digunakan yaitu data jumlah kejadian bencana alam tanah longsor, banjir, banjir bandang, gempa bumi, gelombang pasang laut, angin topan, kebakakaran hutan, dan kekeringan beserta jumlah korban jiwanya tahun 2015 sampai dengan 2017 yang merupakan

peubah numerik. Kemudian peubah kategorik yang terdiri dari informasi topografi wilayah desa, keberadaan pemukiman di puncak atau lereng, keberadaan pemukiman di bantaran sungai, keberadaan wilayah yang berbatasan dengan laut serta lokasi desa terhadap hutan.

### Metode Penelitian

#### *K-Prototypes*

*K-Prototypes* merupakan suatu algoritma analisis *cluster* yang dapat digunakan untuk mengelompokkan objek berdasarkan data campuran yaitu data numerik dan kategorik. Algoritma ini diperkenalkan oleh Huang pada tahun 1998. *K-Prototypes* mendefinisikan sejumlah  $k$  individu virtual atau prototipe sebagai pusat dari kelompok yang dibangun dari mean dari suatu kelompok untuk peubah numerik dan modus untuk peubah kategorik. Pengukuran jarak untuk peubah numerik menggunakan jarak Euclidean sementara untuk peubah kategorik menggunakan jarak *Hamming*.

Pada prakteknya algoritma *K-Prototypes* mirip dengan algoritma *K-Means*, dimana pada awalnya dipilih sejumlah  $k$  prototipe sebagai pusat dari kelompok sementara. Selanjutnya setiap objek dialokasikan kepada prototipe yang terdekat. Ketika semua objek sudah teralokasikan, prototipe diperbaharui kembali untuk menghasilkan jumlah kelompok yang optimal. Kemudian apabila diperlukan objek-objek tersebut dialokasikan ulang ke prototipe yang sudah diperbaharui. Proses tersebut dilakukan secara berulang sampai partisi yang dibuat stabil.

Algoritma ini dibangun di atas tiga proses, pemilihan prototipe awal, alokasi awal, dan realokasi. Ada empat langkah dari algoritma *K-Prototypes*:

- Langkah 1: Tentukan *centroid* dari cluster sebanyak  $k$  cluster sebagai titik awal  $C_1, C_2, \dots, C_k$  pada setiap variabel  $\{X_1, X_2, \dots, X_p\}$ ;
- Langkah 2: Hitung jarak titik data pada dataset terhadap *centroid cluster*, kemudian alokasikan titik data ke *cluster* yang memiliki jarak prototipe terdekat dengan *centroid*;
- Langkah 3: Hitung *centroid* baru *cluster* setelah semua objek telah dialokasikan ke dalam *cluster*, dan kemudian mengalokasikan kembali semua objek pada prototipe baru;
- Langkah 4: Jika *centroid cluster* tidak berubah atau telah konvergen, algoritma akan berhenti. Namun, jika *centroid* masih berubah secara signifikan, proses harus kembali ke langkah 2 dan 3 sampai iterasi maksimum tercapai atau tidak ada pergerakan objek.

#### *Koefisien Silhouette*

Koefisien *silhouette* digunakan untuk melihat kualitas dan kekuatan gerombol, seberapa baik atau buruknya suatu obyek ditempatkan dalam suatu gerombol. Metode ini diusulkan oleh statistikawan asal Belgia, Peter Rousseeuw pada tahun 1987. Metode ini merupakan gabungan dari metode separasi dan kohesi.

Nilai Koefisien *silhouette* adalah ukuran seberapa mirip suatu obyek dengan gerombolnya sendiri (kohesi) dibandingkan dengan gerombol yang lainnya (separasi). Nilai Koefisien *silhouette* berkisar antara -1 hingga +1, dimana nilai tinggi menunjukkan bahwa obyek cocok dengan gerombolnya sendiri dan tidak cocok dengan gerombol tetangga. Jika sebagian besar obyek memiliki nilai yang tinggi, maka konfigurasi penggerombolan sudah sesuai. Jika banyak titik memiliki nilai rendah atau negatif, maka konfigurasi penggerombolan mungkin memiliki terlalu banyak atau terlalu sedikit gerombol.

Untuk menghitung Koefisien *silhouette*, kohesi cluster (a) dan separasi cluster (b) harus dihitung. Kohesi cluster mengacu pada jarak rata-rata antara *instance* (sampel) dan semua titik data lainnya dalam cluster yang sama sementara separasi cluster mengacu pada jarak rata-rata antara *instance* (sampel) dan semua titik data lain di cluster terdekat. Koefisien *silhouette* dapat dihitung seperti yang

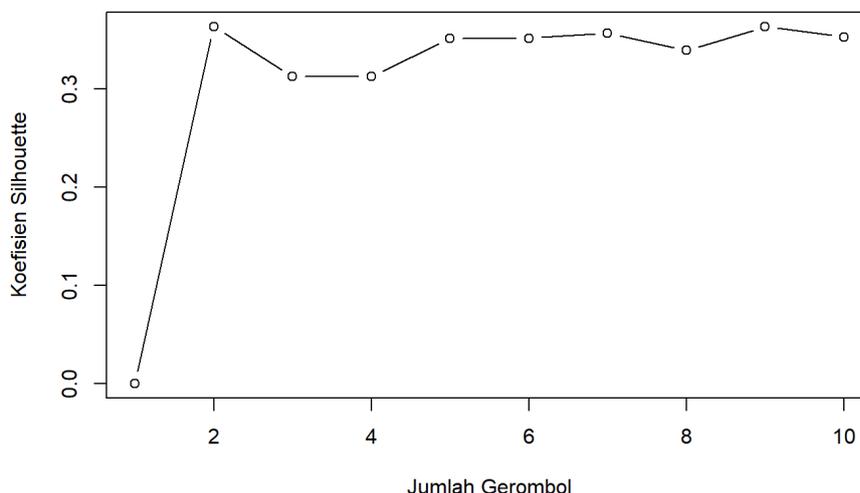
ditunjukkan di bawah ini. Koefisien siluet pada dasarnya adalah perbedaan antara pemisahan cluster dan kohesi dibagi dengan maksimum keduanya (Rousseeuw, 1987).

$$S = \frac{b-a}{\max(a,b)} \tag{1}$$

## HASIL DAN PEMBAHASAN

### Penentuan Jumlah Gerombol

Algoritma *k-prototypes* merupakan salah satu penggerombolan non-hierarki dimana jumlah gerombol harus diketahui dari awal. Pemilihan jumlah gerombol yang optimal dilakukan dengan menggunakan metode koefisien *silhouette*. Gerombol dengan nilai koefisien *silhouette* terbesar merupakan jumlah gerombol yang optimal. Pemilihan gerombol dibatasi dari 2 gerombol sampai dengan 10 gerombol, hal ini dilakukan untuk mempermudah interpretasi karakteristik gerombol. GAMBAR 1 menunjukkan grafik nilai koefisien *silhouette* untuk masing-masing jumlah gerombol. Dari grafik tersebut nilai koefisien *silhouette* terbesar ada pada 9 gerombol dengan nilai 0.3631, sehingga jumlah gerombol yang optimal adalah 9.



GAMBAR 1. Koefisien *silhouette* setiap jumlah gerombol

### Hasil Analisis Cluster

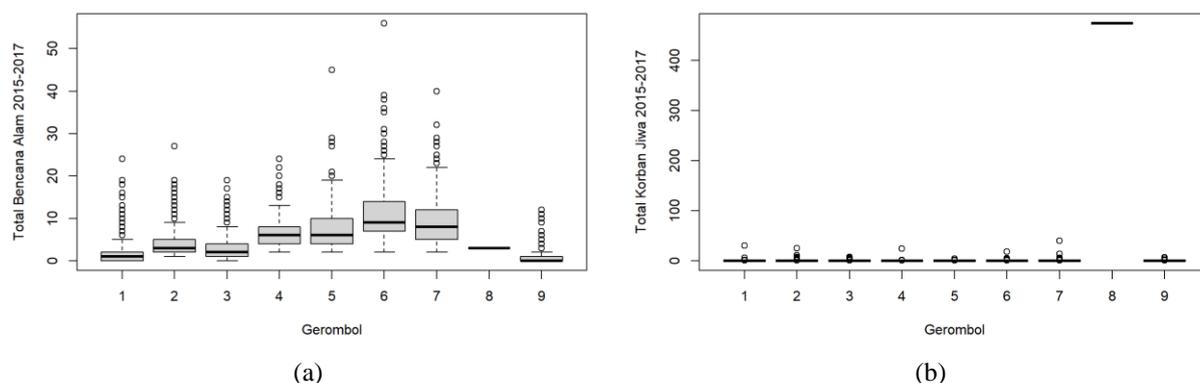
Jumlah desa untuk masing-masing gerombol hasil dari penggerombolan dengan menggunakan algoritma *k-prototypes* dengan jumlah gerombol sebanyak 9 dapat dilihat pada TABEL 1. Berdasarkan tabel tersebut terdapat 2 gerombol dengan jumlah desa cukup ekstrim yaitu gerombol 8 dan 9. Gerombol 9 memiliki jumlah desa yang paling banyak yaitu 2,458 desa, sementara itu gerombol 1 memiliki jumlah desa paling yang sedikit yaitu 1 desa. Selain kedua gerombol tersebut jumlah desa di gerombol lainnya tersebar dari 281 hingga 833 desa.

TABEL 1. Jumlah desa setiap gerombol

| Gerombol | Jumlah Desa |
|----------|-------------|
| 1        | 833         |
| 2        | 665         |
| 3        | 814         |
| 4        | 312         |
| 5        | 303         |
| 6        | 281         |
| 7        | 290         |
| 8        | 1           |
| 9        | 2,458       |

**Karakteristik Gerombol berdasarkan Peubah Numerik**

Karakteristik dari 9 gerombol dapat dilihat berdasarkan informasi dari masing-masing peubah untuk setiap gerombol. Distribusi peubah numerik untuk masing-masing gerombol dapat dilihat pada diagram kotak baris (*boxplot*) dalam GAMBAR 2. Dari GAMBAR 2a dapat dilihat bahwa frekuensi bencana alam paling sering terjadi di gerombol 6 karena memiliki median paling besar. Sementara itu frekuensi bencana alam paling rendah ada pada gerombol 9. Perbedaan frekuensi bencana alam tersebut dapat dikelompokkan menjadi 3 kategori yang berbeda, yaitu tinggi, sedang, dan rendah. Besarnya nilai selang tiga kategori tersebut dihitung dari nilai kuantil yang membagi data frekuensi bencana alam menjadi 3 bagian. Kategori tinggi yaitu frekuensi bencana alam lebih dari 3, kategori sedang dengan frekuensi bencana alam lebih dari 0 hingga 3, sementara rendah yaitu frekuensi bencana alam 0. Dengan kategori tersebut, setiap gerombol dapat dikelompokkan berdasarkan nilai mediannya. Gerombol dengan frekuensi bencana alam tinggi yaitu 4, 5, 6, dan 7. Gerombol dengan frekuensi bencana alam sedang yaitu 1, 2, 3, dan 8. Sementara itu gerombol 9 merupakan satu-satunya yang memiliki frekuensi bencana alam rendah.



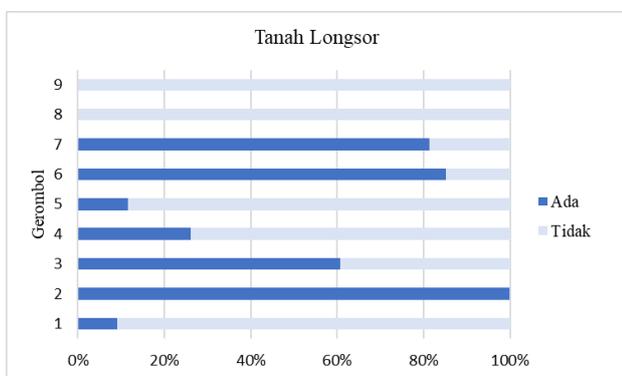
GAMBAR 2. Diagram kotak baris (*boxplot*) frekuensi bencana alam (a) dan total korban jiwa (b) pada tahun 2015-2017 setiap gerombol

GAMBAR 2b menampilkan distribusi setiap gerombol untuk peubah total korban jiwa bencana alam sepanjang 2015 sampai dengan 2017. Dari diagram kotak baris tersebut dapat dilihat bahwa gerombol 8 memiliki total korban jiwa yang paling ekstrim dengan median 474 dibandingkan cluster lainnya yang memiliki median 0. Dari data tersebut dapat dikategorikan bahwa gerombol 8 memiliki tingkat korban jiwa bencana alam yang tinggi sementara cluster lain rendah. Gerombol 8 terdiri dari 1 desa yaitu Desa Kota Baru yang terletak di Kecamatan Bekasi Barat, Kota Bekasi. Total 474 korban jiwa pada gerombol ini merupakan akumulasi korban jiwa bencana banjir pada tahun 2015 sebanyak 147 jiwa, tahun 2016 sebanyak 158 jiwa, dan tahun 2017 sebanyak 169 jiwa.

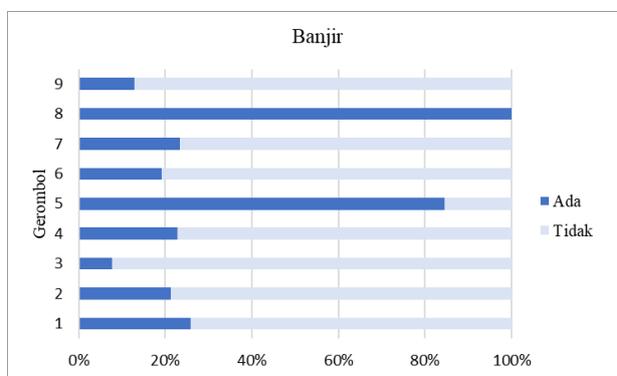
**Karakteristik Gerombol berdasarkan Peubah Kategorik**

Berdasarkan peubah kategorik, gerombol yang terbentuk dapat dilihat karakteristiknya berdasarkan jenis bencana yang paling mungkin terjadi (GAMBAR 3) dan keadaan wilayah desanya (GAMBAR 4). Berdasarkan GAMBAR 3, bencana alam yang dapat menjadi karakteristik sebuah gerombol yaitu tanah longsor, banjir, gempa bumi, angin topan dan kekeringan karena setidaknya memiliki 1 gerombol dengan persentase lebih dari 50%. Sementara itu bencana alam banjir bandang, gelombang pasang, dan kebakaran hutan tidak memiliki persentase yang dominan di setiap gerombol sehingga tidak dapat dijadikan sebagai karakteristik.

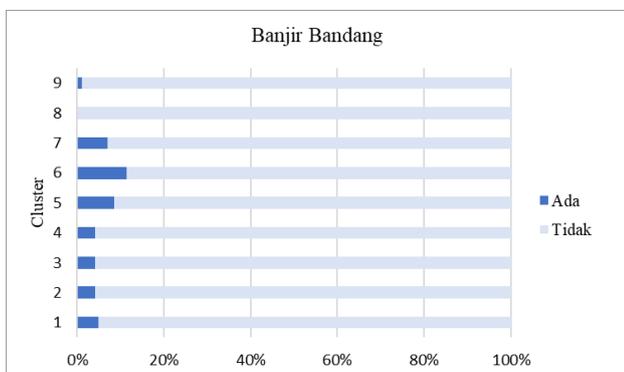
Kejadian tanah longsor memiliki persentase yang besar (lebih dari 50%) di gerombol 2, 3, 6 dan 7. Kejadian bencana banjir memiliki peluang yang tinggi di gerombol 5 dan 8. Kejadian gempa bumi memiliki persentase yang besar di gerombol 4, 6, dan 7. Kejadian angin topan memiliki peluang yang tinggi di gerombol 5 dan 7. Sementara itu kejadian kekeringan memiliki persentase yang besar pada gerombol 4, 5, dan 6.



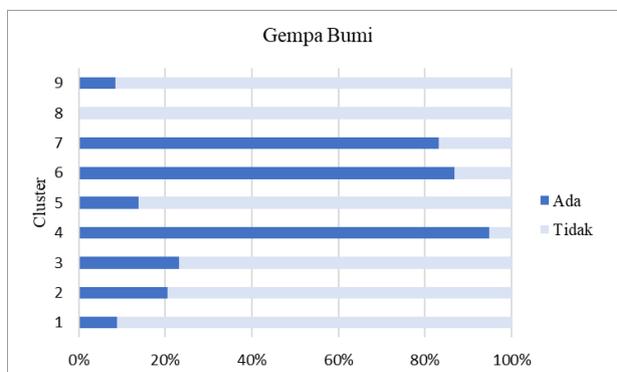
(a)



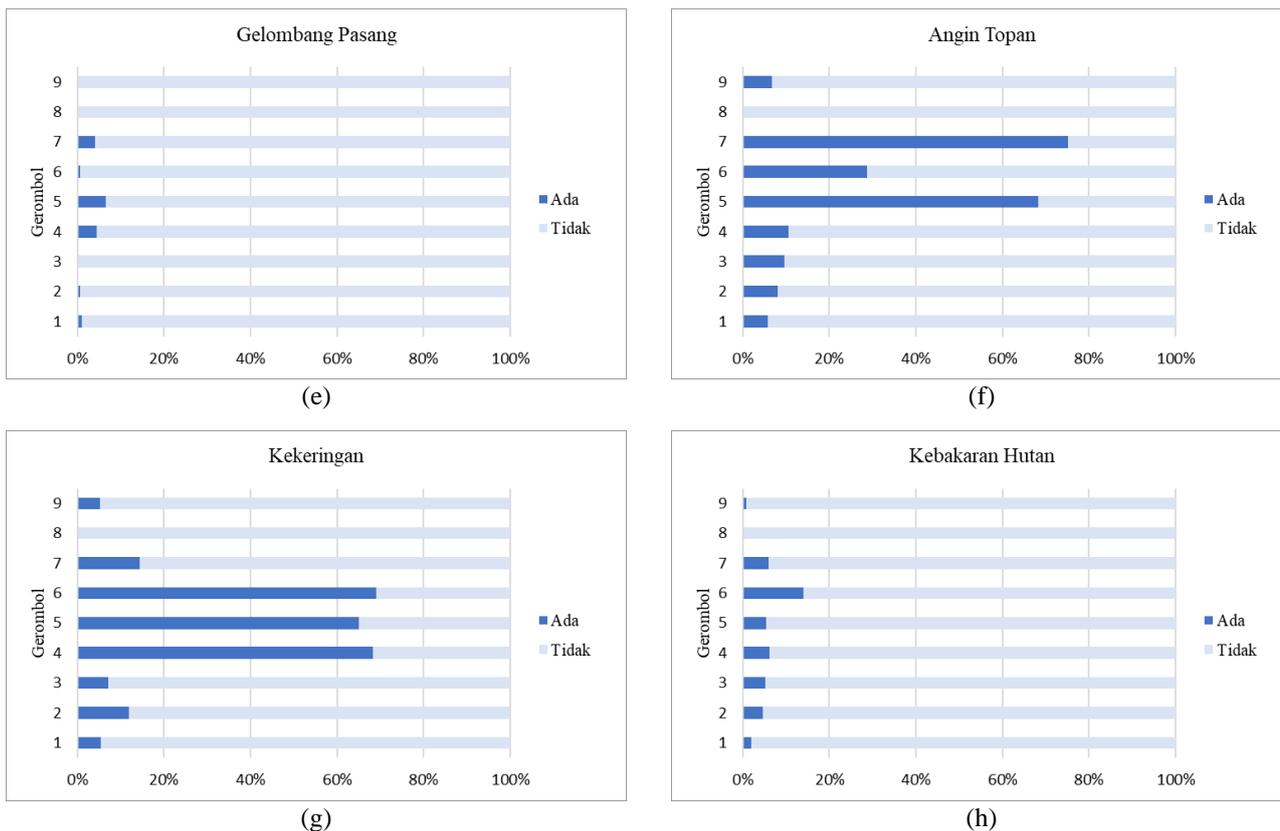
(b)



(c)

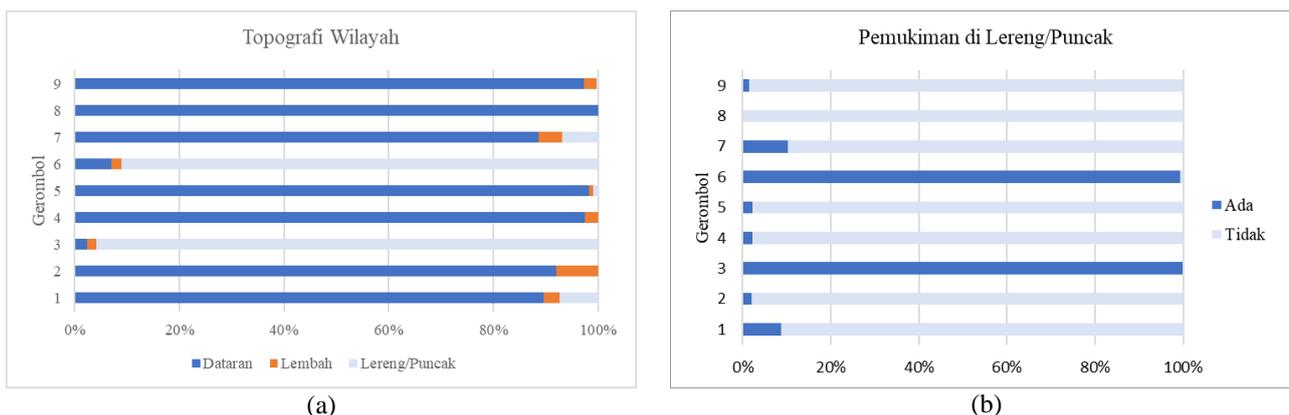


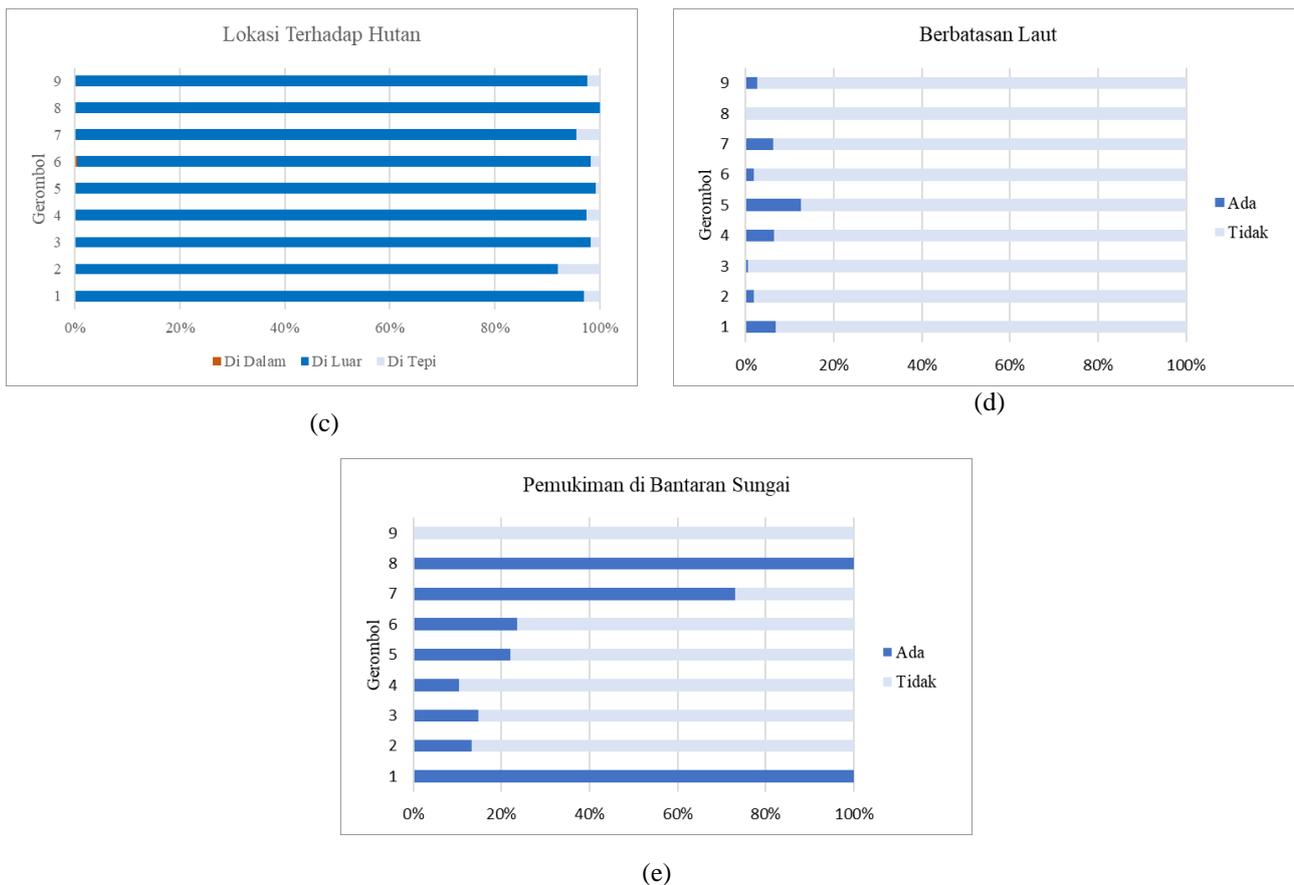
(d)



GAMBAR 3. Persentase berbagai kejadian bencana alam setiap gerombol

Berdasarkan GAMBAR 4, karakteristik peubah yang dapat digunakan adalah topografi wilayah, keberadaan pemukiman di lereng atau puncak dan keberadaan pemukiman di bantaran sungai. Sementara itu lokasi terhadap hutan tidak dapat menjadi karakteristik khusus karena mayoritas semua gerombol berada di luar hutan, begitu pun dengan wilayah berbatasan dengan laut yang mayoritas semua gerombol tidak berbatasan dengan laut.





GAMBAR 4. Karakteristik wilayah setiap gerombol

Secara topografi wilayah gerombol dapat dibedakan menjadi dua, yaitu gerombol dengan topografi lereng atau puncak yang diwakili oleh gerombol 3 dan 6 serta sisa gerombol lainnya dengan topografi dataran. Sejalan dengan topografi wilayah, keberadaan pemukiman di lereng atau puncak pun diwakili oleh gerombol 3 dan 6. Sementara itu untuk pemukiman di bantaran sungai memiliki persentase yang besar di gerombol 1, 7, dan 8.

**KESIMPULAN DAN SARAN**

Penggerombolan desa di Jawa Barat berdasarkan daerah rawan bencana menggunakan algoritma *k-prototypes* menghasilkan 9 gerombol optimal. Berikut ini karakteristik gerombol yang dihasilkan :

TABEL 2. Karakteristik gerombol dan jumlah desa

| Frekuensi Bencana | Tingkat Korban Jiwa | Cluster | Jenis Bencana                | Karakteristik Wilayah                                 | Jumlah Desa |
|-------------------|---------------------|---------|------------------------------|---|-------------|
| Tinggi            | Rendah              | 4       | Gempa bumi dan kekeringan    | Dataran   | 312         |
|                   |                     | 5       | Banjir dan angin topan       | Dataran   | 303         |
|                   |                     | 6       | Tanah longsor dan gempa bumi | Lereng/puncak dan terdapat pemukiman di lereng/puncak | 281         |

|        |        |   |                            |   |       |
|--------|--------|---|----------------------------|---|-------|
|        |        | 7 | Gempa bumi dan angin topan | Dataran, terdapat pemukiman di bantaran sungai        | 290   |
|        | Tinggi | 8 | Banjir                     | Dataran, terdapat pemukiman di bantaran sungai        | 1     |
| Sedang |        | 1 | Banjir                     | Dataran, terdapat pemukiman di bantaran sungai        | 833   |
|        | Rendah | 2 | Tanah longsor              | Dataran   | 665   |
|        |        | 3 | Tanah longsor              | Lereng/puncak dan terdapat pemukiman di lereng/puncak | 814   |
| Rendah | Rendah | 9 |                            | Dataran   | 2,458 |

Penelitian ini masih dapat dikembangkan dengan penambahan peubah yang digunakan baik dari data Podes maupun dari sumber data lainnya. Selain itu, dapat juga dikembangkan dengan menggunakan metode penggerombolan lain untuk data campuran seperti algoritma Kamila dan *Latent Class Model* (LCM).

#### REFERENSI

- Adi, A.W., Shalih, O., Shabrina, F.Z., Rizqi, A., Putra, A.S., Karimah, R., et al, 2022, *Indeks Risiko Bencana Indoensia Tahun 2021*, Pusat Data Informasi dan Komunikasi Kebencanaan Badan Nasional Penanggulangan Bencana, Jakarta.
- BNPB, 2022, *Geoportal Data Bencana Indonesia*, dilihat 4 September 2022, <https://gis.bnpb.go.id/>.
- Faizana, F., Nugraha, A.L. and Yuwono, B.D., 2015. Pemetaan risiko bencana tanah longsor Kota Semarang. *Jurnal Geodesi Undip*, 4(1), pp.223-234.
- Huang, Z., 1998. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data mining and knowledge discovery*, 2(3), pp.283-304.
- Open Data Jabar, 2021, *Bencana Alam di Jawa Barat*, dilihat 4 September 2022, <https://opendata.jabarprov.go.id/id/visualisasi/bencana-alam-di-jawa-barat>.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, pp.53-65.
- Sadewo, M.G., Windarto, A.P. and Wanto, A., 2018. Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 2(1).
- Tondobala, L., 2012. Pemahaman tentang kawasan rawan bencana dan tinjauan terhadap kebijakan dan peraturan terkait. *Sabua: Jurnal Lingkungan Binaan dan Arsitektur*, 3(1).

Triyanto, S., Sunyoto, A. and Arief, M.R., 2021. Analisis klasifikasi bencana banjir berdasarkan curah hujan menggunakan algoritma Naïve Bayes. *JOISIE (Journal of Information Systems and Informatics Engineering)*, 5(2), pp.109-117.