

CLUSTERING KEJADIAN BENCANA ALAM di JAWA BARAT BERDASARKAN JENIS BENCANA MENGGUNAKAN K-MEANS

Indah Rosalياهو¹⁾, Bani Nurhakim²⁾

^{1,2)} Jurusan Manajemen Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No 10B, Kota Cirebon 45131, Jawa Barat, Indonesia
E-mail: ¹⁾rosalياهوindah@gmail.com, ²⁾baninurhakim@gmail.com

ABSTRAK

Jawa barat merupakan salah satu wilayah dengan potensi bencana alam tinggi. Hampir semua jenis bencana sudah terjadi di setiap wilayahnya, seperti gempa bumi, tanah longsor, banjir, dan masih banyak lagi. Oleh karena itu, informasi mengenai tingkat terjadinya bencana alam di berbagai wilayah harus diteliti lebih lanjut agar lebih waspada kedepannya. Penelitian ini bertujuan untuk mengelompokkan kejadian bencana alam di Jawa Barat berdasarkan jenis bencana dengan memanfaatkan teknik *clustering* pada *data mining*. Proses pengelompokan data dilakukan menggunakan metode algoritma *K-Means* dan tahap perancangan yang digunakan yaitu *Knowledge Discovery in Database* (KDD). Dengan menggunakan *tools RapidMiner* diperoleh 6 *cluster* dengan nilai *Davies Bouldin Index* yaitu 9.20. *Cluster 3* merupakan daerah dengan kejadian bencana alam sangat rendah, *cluster 1* daerah dengan kejadian bencana alam rendah, *cluster 4* daerah dengan kejadian bencana alam sedang, *cluster 0* daerah dengan kejadian bencana alam tinggi 1, *cluster 5* daerah dengan kejadian bencana alam tinggi 2, dan *cluster 2* merupakan daerah dengan kejadian bencana alam sangat tinggi.

Kata kunci : bencana alam, *clustering*, *means*

ABSTRACT

West Java is one of the areas with high potential for natural disasters. Almost all types of disasters have occurred in each region, such as earthquakes, landslides, floods, and many more. Therefore, information about the level of occurrence of natural disasters in various regions must be studied further in order to be more vigilant in the future. This research aims to group natural disaster events in West Java based on the type of disaster by utilizing clustering techniques in data mining. The data clustering process is carried out using the K-Means algorithm method and the design stage used is Knowledge Discovery in Database (KDD). By using RapidMiner tools, 6 clusters were obtained with a Davies Bouldin Index value of 9.20. Cluster 3 is an area with very low natural disaster occurrence, cluster 1 is an area with low natural disaster occurrence, cluster 4 is an area with moderate natural disaster occurrence, cluster 0 is an area with high natural disaster occurrence 1, cluster 5 is an area with high natural disaster occurrence 2, and cluster 2 is an area with very high natural disaster occurrence.

Keywords : natural disaster, *clustering*, *means*

1. PENDAHULUAN

Indonesia adalah salah satu wilayah dengan potensi kejadian bencana alam yang tinggi di

Asia Tenggara[1]. Letak geologis Indonesia menyebabkan Indonesia banyak mengalami kejadian bencana alam[2]. Jawa Barat merupakan salah satu provinsi yang rawan

bencana alam. Hampir semua jenis bencana sudah terjadi di setiap wilayahnya, seperti gempa bumi, tanah longsor, banjir, angin puting beliung dan masih banyak lagi. Badan Penanggulangan Bencana Daerah (BPBD) mencatat sekitar 2.400 kejadian bencana alam terjadi sepanjang tahun 2021[3].

Teknologi yang semakin canggih mampu menjelaskan gejala awal bencana, namun kejadian detil dari bencana alam tersebut hanya bisa diprediksi. Meskipun begitu, dengan mengetahui informasi terkait intensitas kejadian bencana alam di setiap daerah, maka manusia dapat mempersiapkan diri dalam menanggulangi bencana alam kedepannya. upaya mitigasi tersebut diantaranya memfasilitasi ketersediaan tempat tinggal, kemungkinan rehabilitasi dan sumber daya alam yang ada [4].

Semakin banyak bencana alam yang terjadi, maka semakin banyak pula kerugian yang akan ditimbulkan. Oleh karena itu, kejadian bencana alam di Jawa Barat perlu dikelompokkan berdasarkan jenis bencana agar dapat mengetahui intensitas kejadian bencana alam di setiap daerah. Mengingat semakin banyaknya data yang ada, maka pendekatan yang dilakukan yaitu menggunakan teknik *clustering* pada *data mining*. Metode yang akan digunakan pada penelitian ini yaitu *K-Means*. Algoritma *K-Means* yaitu metode untuk mengelompokkan suatu data berdasarkan pencarian pusat *cluster* secara iterasi dimana letak *cluster* setiap data dihasilkan jarak minimum pusat *cluster* [5]. Metode algoritma *K-Means* bukan hal baru yang digunakan dalam pengelompokan, hal ini dibuktikan dengan adanya penelitian terdahulu yang menggunakan teknik *clustering* dengan algoritma *K-Means*. penelitian yang dilakukan oleh Devy Isya Ramadhani et.al (2022) membahas tentang penerapan metode *K-Means* untuk *clustering* desa rawan bencana berdasarkan data kejadian terjadinya bencana alam [1].

Berdasarkan masalah yang sudah di paparkan, bencana alam di Jawa Barat terjadi setiap tahunnya. Oleh karena itu, pemerintah provinsi Jawa Barat harus lebih siap lagi dalam

menghadapi bencana alam yang kemungkinan akan terjadi di kemudian hari, sehingga diperlukan adanya pengelompokan kejadian bencana alam berdasarkan jenis bencana agar dapat memberikan informasi wilayah mana saja yang sering mengalami kejadian bencana alam di Jawa Barat.

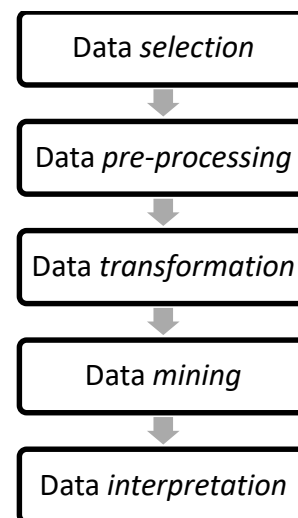
2. METODE PENELITIAN

2.1. Pengumpulan Data

Teknik pengumpulan data pada penelitian ini menggunakan data sekunder. Dimana data bersumber dari *website* Open Data Jabar <https://opendata.jabarprov.go.id/> data tentang kejadian bencana di Jawa Barat dari tahun 2012-2021 memuat 12 atribut dan 270 *record*.

2.2. Perancangan Data

Tahapan perancangan pada penelitian ini menggunakan tahapan *Knowledge Discovery In Databases* (KDD). Tahapan dalam KDD antara lain:



Gambar 1. Tahapan KDD

1. Data Selection

Data selection merupakan proses menyeleksi data yang digunakan sehingga sesuai dengan tujuan penelitian [6]. Dari 12 atribut akan di seleksi menjadi 8 atribut saja yang diperlukan yaitu nama kabupaten/kota, banjir, tanah longsor, gempa bumi, puting beliung, gelombang pasang, satuan dan tahun.

2. Data Pre-processing/Cleaning

Dataset yang tidak mempunyai nilai (*missing*) akan dihilangkan. Sehingga, hasilnya adalah *dataset* matang yang akan di proses dalam tahap selanjutnya.

3. Data Transformation

Suatu proses penyesuaian data dengan cara menggabungkan atau mengubah data sesuai dengan format yang akan diproses ke *data mining* [7]. Pada tahap ini, data dengan tipe nominal seperti nama kabupaten akan diubah ke bentuk *numerical*. Tujuannya adalah agar data dapat di proses menggunakan algoritma *K-Means*.

4. Data Mining

Data akan di proses menggunakan algoritma *K-Means*, jumlah pengelompokan akan ditentukan terlebih dahulu sehingga mencapai nilai K yang paling optimum. Proses *data mining* terhadap bencana alam berdasarkan jenis bencana akan menggunakan *tools RapidMiner*.

Tahapan algoritma *K-Means* yaitu:

- Menentukan nilai K secara acak [8].
- Menentukan titik pusat (*centroid*) awal secara acak [9].
- Menghitung jarak setiap data terhadap *centroid* menggunakan rumus jarak *Euclidean Distance*. Berikut merupakan rumus *Euclidean Distance*:

$$d(x, \mu) = \sqrt{\sum(x - \mu)^2} \quad (1)[10]$$

5. Data Interpretation

Pada tahap ini, data yang sudah di proses akan di evaluasi menggunakan model *performance* untuk menghitung *Davies Bouldin Index* (DBI), dengan tujuan memperoleh nilai K yang paling optimal. Rumus DBI adalah sebagai berikut:

$$DBI = \frac{1}{c} \sum_{i=1}^c Ri \quad (2)$$

dengan

$$Ri = \max_{j=1, \dots, C, i \neq j} Rij, \quad Rij = \frac{Si + Sj}{d(Vi, Vj)} \quad (3)$$

[11]

dimana:

C : Jumlah *cluster*

Rij : Ukuran kemiripan *cluster* ke-i dan ke-j

Si : Ukuran dispersi *cluster* i, i=1,2,3...C

3. HASIL DAN DISKUSI

Dataset kejadian bencana alam di Jawa Barat sudah dalam bentuk *excel*, dimana data sudah masuk kedalam proses *pre-processing* dan proses *transformation* data. Proses selanjutnya yaitu *data mining* dengan algoritma *K-Means*.

Tabel 1. *Dataset* Kejadian Bencana Alam Tahun 2012-2021

NKK	B	TL	GB	PB	GP	T
Kabupaten Bogor	0	0	3	2	0	2012
Kabupaten Sukabumi	20	48	41	27	0	2012
Kabupaten Cianjur	1	9	0	23	0	2012
Kabupaten Bandung	17	12	0	3	0	2012
Kabupaten Garut	7	36	0	28	1	2012
Kabupaten Tasikmalaya	0	2	0	1	0	2012
Kabupaten Ciamis	0	2	0	2	0	2012
Kabupaten Kuningan	6	17	0	2	0	2012
Kabupaten Cirebon	0	0	0	1	0	2012
Kabupaten Majalengka	0	0	0	6	0	2012
...						
Kota Tasikmalaya	2	2	0	7	0	2021
Kota Banjar	1	3	1	9	0	2021

Keterangan :

NKK : Nama kabupaten/kota

B : Banjir

TL : Tanah longsor

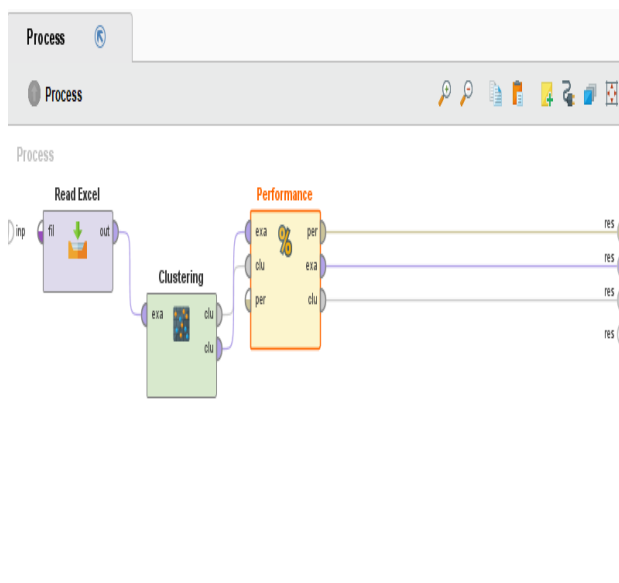
GB : Gempa bumi

PB : Puting beliung

GP : Gelombang pasang

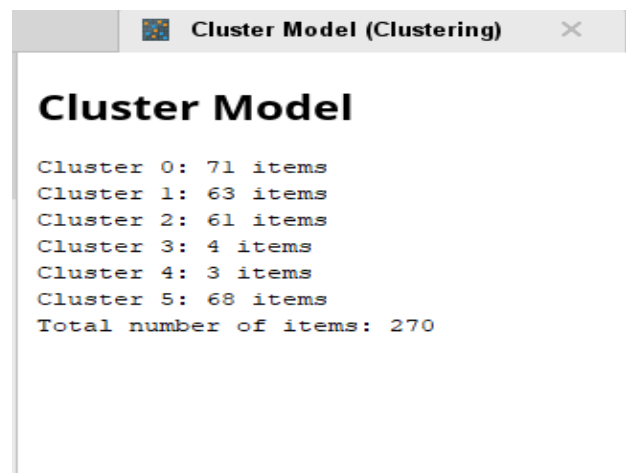
T : Tahun

Dalam penelitian ini, data akan di olah menggunakan *tools RapidMiner*. Proses dapat dilihat pada Gambar 2, dimana terdapat 3 operator. Yaitu *read excel* yang berfungsi untuk membaca data dalam format *excel*, *clustering* merupakan model algoritma untuk melakukan proses data *mining* dan operator untuk mengetahui nilai DBI (*Davies Bouldin Index*) yaitu *cluster distance performance*.



Gambar 2. Proses *Clustering* Menggunakan *K-Means*

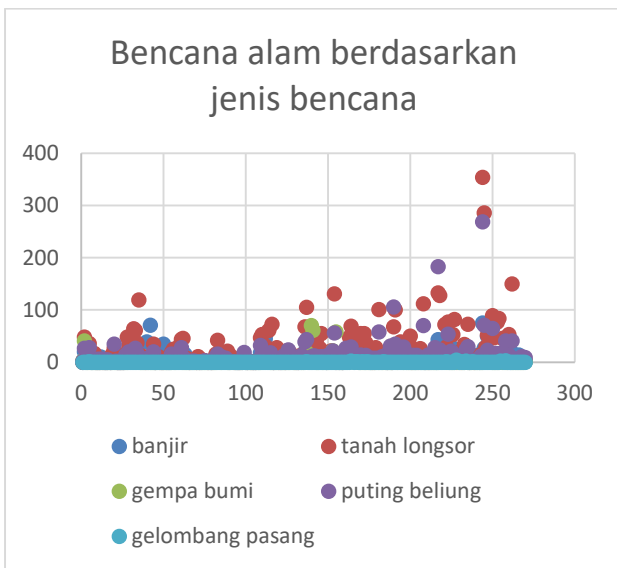
Setelah dilakukan proses *data mining* menggunakan algoritma *K-Means* menghasilkan 6 *cluster* dari 270 *record* data. *Cluster* 0 dengan jumlah anggotanya adalah sebanyak 71 kabupaten/kota, *cluster* 1 sebanyak 63 kabupaten/kota, *cluster* 2 sebanyak 61 kabupaten/kota, *cluster* 3 sebanyak 4 kabupaten/kota, *cluster* 4 merupakan yang paling sedikit yaitu 3 kabupaten/kota, dan yang terakhir adalah *cluster* 5 dengan jumlah keanggotaannya sebanyak 68 kabupaten/kota, sehingga jumlah secara keseluruhan adalah 270 kabupaten/kota dari tahun 2012 – 2021.



Gambar 3. Hasil *Cluster Model*

Setelah melakukan pemodelan pada pengelompokan data bencana, selanjutnya adalah melakukan validasi *performance* untuk mengetahui nilai DBI (*Davies Bouldin Index*). Dengan menggunakan *cluster distance performance* pada nilai $K=6$ diperoleh hasil DBI nya adalah 0.920.

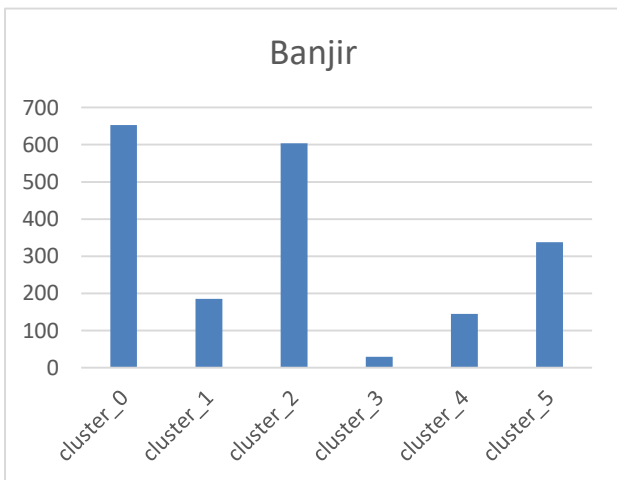
Berdasarkan penyebaran kejadian bencana alam dari tahun 2012 – 2021 jenis bencana yang paling sering terjadi adalah tanah longsor. Dari Gambar 4 tanah longsor menempati posisi paling tinggi yaitu mencapai 354 kejadian di kabupaten Bogor. Puting beliung merupakan paling tinggi tingkat kejadian bencana kedua setelah tanah longsor, dengan kejadian bencana mencapai 269 terjadi di kabupaten Bogor. Bencana alam yang sering terjadi berikutnya adalah banjir, dengan tingkat kejadian bencana mencapai 75 terjadi di kabupaten Bogor. Gempa bumi merupakan kejadian bencana alam sedang, dengan banyaknya kejadian mencapai 70 di kabupaten Garut. Bencana alam yang paling sedikit kejadiannya adalah gelombang pasang, dengan kejadian bencana hanya mencapai 4 kejadian di kabupaten Indramayu.



Gambar 4. Scatter Plot Bencana Alam Berdasarkan Jenis Bencana

Gambar 5 sampai dengan Gambar 9 merupakan gambar hasil *cluster* pada masing-masing atribut kejadian bencana alam yaitu banjir, tanah longsor, gempa bumi, puting beliung dan gelombang pasang.

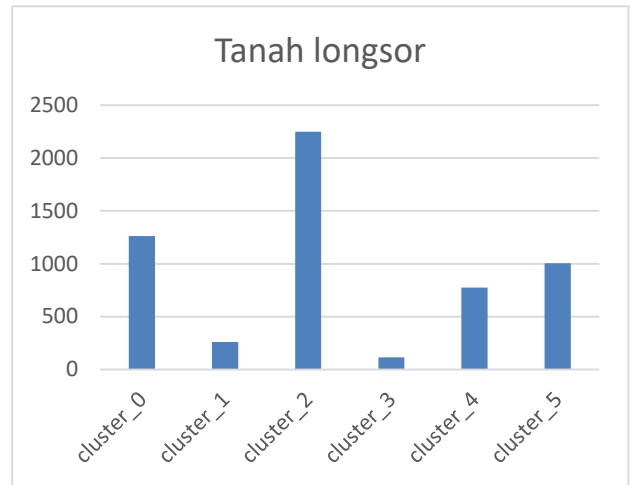
Banjir paling banyak terjadi pada *cluster* 0, artinya *cluster* 0 merupakan daerah dengan kejadian bencana alam yang paling mendominasi adalah banjir. Sedangkan yang paling sedikit terjadinya banjir terdapat pada *cluster* 3.



Gambar 5. Cluster Terhadap Banjir

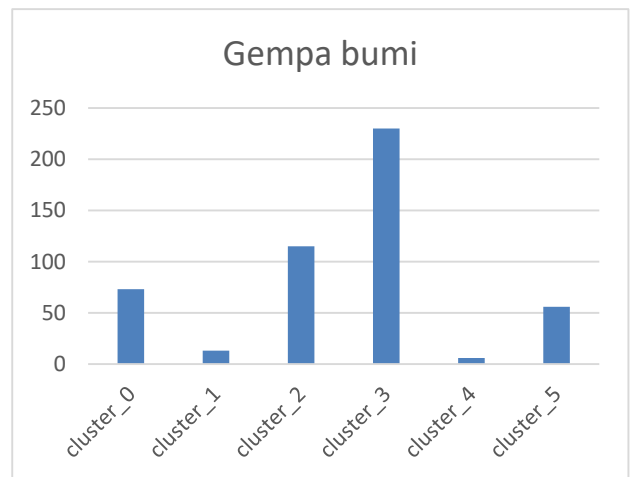
Cluster 2 merupakan daerah dengan bencana alam yang paling mendominasi adalah tanah

longsor, sedangkan yang paling rendah terjadi di *cluster* 3.



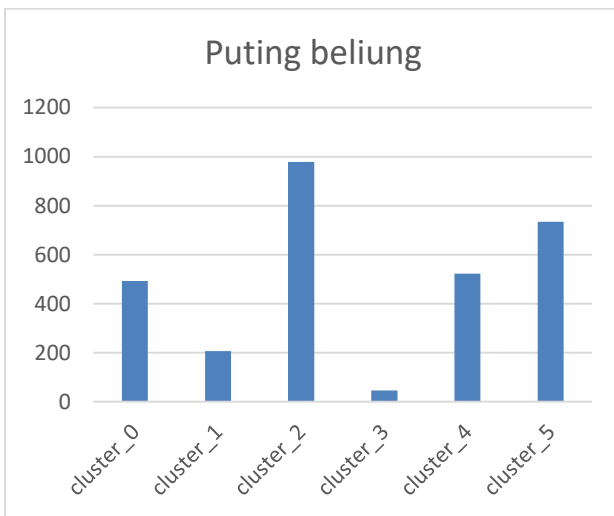
Gambar 6. Cluster terhadap Tanah longsor

Gempa bumi paling banyak terjadi pada *cluster* 3, yang paling rendah terdapat pada *cluster* 4 dan *cluster* 1, dan yang sedang kejadiannya terdapat pada *cluster* 0, *cluster* 5 dan *cluster* 2.



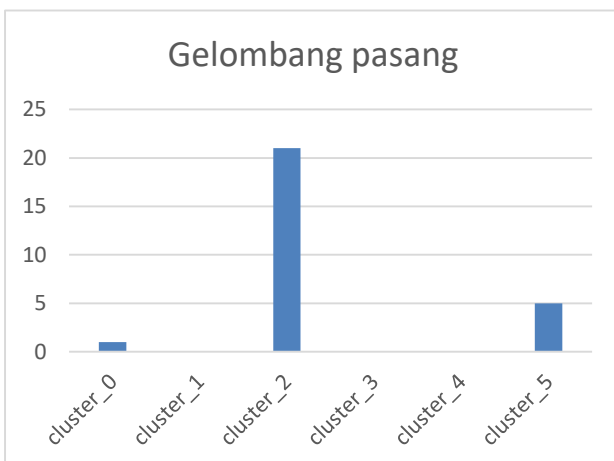
Gambar 7. Cluster Terhadap Gempa bumi

Puting beliung paling banyak terjadi pada *cluster* 2. *Cluster* 0 dan *cluster* 4 menempati posisi yang hampir sama, sedangkan *cluster* 3 merupakan yang paling sedikit kejadian puting beliung.



Gambar 8. Cluster Terhadap Puting beliung

Gelombang pasang merupakan jenis bencana alam yang paling sedikit kejadiannya dari tahun 2012 – 2021. Pada *cluster 1*, *cluster 3* dan *cluster 4* tidak ada jenis bencana gelombang pasang. Sedangkan yang paling tinggi kejadiannya adalah pada *cluster 2*. Hal ini berarti gelombang pasang hanya terdapat pada beberapa cluster saja yaitu *cluster 2*, *cluster 5* dan *cluster 0*.



Gambar 9. Cluster terhadap Gelombang pasang

4. KESIMPULAN DAN SARAN

Dari hasil *clustering* menggunakan algoritma *K-Means*, data kejadian bencana alam berdasarkan jenis bencana dapat dikelompokkan menjadi 6 *cluster*. *Cluster 3* : sangat rendah yaitu 343 kejadian, *cluster 1* :

rendah dengan 405 kejadian, *cluster 4*: sedang dengan 924 kejadian, *cluster 0* : tinggi 1 dengan 1337 kejadian, *cluster 5*: tinggi 2 dengan 1403 kejadian, dan *Cluster 2*: sangat tinggi yaitu 2269 kejadian.

Peneliti selanjutnya diharapkan dapat menambahkan atribut pendukung dalam mengetahui daerah yang rawan dan tidak rawan bencana alamnya.

5. DAFTAR PUSTAKA

- [1] D. I. Ramadhani, O. Damayanti, O. Thaushiyah, and A. R. Kadafi, "Penerapan Metode K-Means Untuk Clustering Desa Rawan Bencana Berdasarkan Data Kejadian Terjadinya Bencana Alam," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, pp. 749–753, Jun. 2022, Accessed: May 06, 2023. [Online]. Available: <http://dx.doi.org/10.30865/jurikom.v9i3.4326>
- [2] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, pp. 744–752, 2020.
- [3] BPBD Provinsi Jawa Barat, "Longsor Dominasi Bencana Alam di Jabar," <https://bpbj.jabarprov.go.id/>, Jan. 03, 2022. <https://bpbj.jabarprov.go.id/longsor-dominasi-bencana-alam-di-jabar/> (accessed May 07, 2023).
- [4] Y. Pratama, H. Hendrawan, E. Rasywir, B. T. Carenina, and D. R. Anggraini, "Penerapan Algoritma K-Means clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1232–

- 1240, 2022, Accessed: May 07, 2023. [Online]. Available: doi:10.47065/bits.v4i3.2549
- [5] F. M. Basysyar, “Clustering Data Disabilitas menggunakan Algoritma K-Means di Kabupaten Cirebon,” *JURSIMA (Jurnal Sistem Informasi dan Manajemen)*, vol. 9, no. 3, pp. 247–255, 2021, Accessed: May 07, 2023. [Online]. Available: <https://ejournal.stmikgici.ac.id/index.php/jursima/article/view/305>
- [6] S. C. Nurzanah, S. Alam, and T. I. Hermanto, “Analisis Association Rule Untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori (Studi Kasus: Klinik Rafina Medical Center),” *JIKO (Jurnal Informatika dan Komputer)*, vol. 5, no. 2, pp. 132–141, 2022, doi: 10.33387/jiko.v5i2.4792.
- [7] A. H. Ardiansyah, W. Nugroho, N. H. Alfiah, R. A. Handoko, and M. A. Bakhtiar, “Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan Status Provinsi di Indonesia 2020,” in *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, 2020, pp. 329–333. doi: 10.29407/inotek.v4i3.108.
- [8] M. A. Putri, N. Rahaningsih, F. M. Basysyar, and O. Nurdiawan, “Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan,” *Jurnal Accounting Information System (AIMS)*, vol. 5, no. 2, pp. 145–156, 2022.
- [9] H. Havaluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, “Implementasi Metode K-Means Untuk Pengelompokan Rekomendasi Tugas Akhir,” *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, vol. 16, no. 1, pp. 13–18, 2021, doi: <http://dx.doi.org/10.30872/jim.v16i1.5182>.
- [10] S. Handoko, F. Fauziah, and E. T. E. Handayani, “Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering,” *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: <http://dx.doi.org/10.35760/tr.2020.v25i1.2677>.
- [11] A. Badruttamam, S. Sudarno, and I. M. Di Asih, “Penerapan Analisis Kluster K-Modes Dengan Validasi Davies Bouldin Index Dalam Menentukan Karakteristik Kanal Youtube di Indonesia (Studi Kasus: 250 Kanal Youtube Indonesia Teratas Menurut Socialblade),” *Jurnal Gaussian*, vol. 9, no. 3, pp. 263–272, 2020, doi: <https://doi.org/10.14710/j.gauss.9.3.263-272>.