

IDENTIFIKASI KEPADATAN PENDUDUK DI PROVINSI JAWA BARAT MENGGUNAKAN HIERARCHICAL CLUSTERING

Wahyu Iskandar Simanjuntak^{1)*}, Yusran Timur Samuel²⁾

^{1,2}Jurusan Teknik Informatika

^{1,2}Fakultas Teknik Informasi

^{1,2}Universitas Advent Indonesia

E-mail : 2181001@unai.edu^{1)*}, y.tarihoran@unai.edu²⁾

Abstract

This research applies a hierarchical clustering algorithm to identify population density patterns in West Java (18 regencies, 9 cities) as a basis for natural disaster management. Population density data for 2020-2022 from the West Java Population Office was analyzed to group areas into three categories: densest, medium, and lowest. The hierarchical clustering method was used to group areas based on population density and flood potential, with the additional attribute of river presence. The clustering results were evaluated using the Davies-Bouldin index. The results showed that the algorithm was successfully applied, grouping 20 districts/cities with the lowest population density (Cluster 0), 3 districts/cities with medium density (Cluster 1), and 4 districts/cities with the densest density (Cluster 2). This research is expected to provide insight to the government and related institutions in planning disaster mitigation based on population density patterns, so as to reduce the risk of natural disasters in the future. This research takes data from the official source <https://jabar.bps.go.id/indicator/12/245/1/kepadatan-penduduk-menurut-kabupaten-kota.html>. The main objective of this research is to understand population density patterns that can provide an indication of the level of risk to certain natural disasters in the West Java region. This information is expected to be used as a basis for more effective and efficient disaster management strategies in the future. The implication of this research shows that by understanding the pattern of population density and river distribution through the hierarchical clustering method, the government and related institutions can formulate more targeted disaster management strategies.

Keywords- Hierarchical Clustering, Population density, Disaster mitigation, river

Intisari

Penelitian ini menerapkan algoritma hierarchical clustering untuk mengidentifikasi pola kepadatan penduduk di Jawa Barat (18 kabupaten, 9 kota) sebagai dasar penanggulangan bencana alam. Data kepadatan penduduk 2020-2022 dari Dinas Kependudukan Jawa Barat dianalisis untuk mengelompokkan wilayah menjadi tiga kategori: terpadat, sedang, dan terendah. Metode hierarchical clustering digunakan untuk mengelompokkan wilayah berdasarkan kepadatan penduduk dan potensi banjir, dengan atribut tambahan yaitu keberadaan sungai. Hasil pengelompokan dievaluasi menggunakan Davies-Bouldin index. Hasilnya menunjukkan bahwa algoritma berhasil diterapkan, mengelompokkan 20 kabupaten/kota dengan kepadatan penduduk terendah (Cluster 0), 3 kabupaten/kota dengan kepadatan sedang (Cluster 1), dan 4 kabupaten/kota dengan kepadatan terpadat (Cluster 2). Penelitian ini diharapkan dapat memberikan wawasan kepada pemerintah dan lembaga terkait dalam merencanakan mitigasi bencana berdasarkan pola kepadatan penduduk, sehingga dapat mengurangi risiko bencana alam di masa depan. Penelitian ini mengambil data dari sumber resmi <https://jabar.bps.go.id/indicator/12/245/1/kepadatan-penduduk-menurut-kabupaten-kota.html>. Tujuan utama penelitian ini adalah untuk memahami pola kepadatan penduduk yang dapat memberikan indikasi tingkat risiko terhadap bencana alam tertentu di wilayah Jawa Barat. Informasi ini diharapkan dapat digunakan sebagai dasar untuk strategi penanggulangan bencana yang lebih efektif dan efisien di masa depan. Implikasi dari penelitian ini menunjukkan bahwa dengan memahami pola kepadatan penduduk dan distribusi sungai melalui metode hierarchical clustering, pemerintah dan lembaga terkait dapat merumuskan strategi penanggulangan bencana yang lebih tepat sasaran.

Kata Kunci— Hierarchical Clustering, Kepadatan penduduk, Penanggulangan bencana, sungai

1. PENDAHULUAN

Banyaknya penduduk bisa sebagai potensi ataupun sebagai kelemahan sebuah negara, dalam negara akan memiliki potensi jika banyak

penduduknya merata dan diiringi dengan kualitas kehidupan yang baik. Namun dapat menjadi kelemahan bilamana banyak penduduk melebihi daya tampung wilayah negara [1]. Tingginya jumlah penduduk disuatu daerah dikarenakan mempunyai daya tarik seperti dari

segi pendidikan, ekonomi, dan wisata [2]. Kepadatan penduduk tidak selalu memiliki pengaruh buruk jika SDM dan infrastruktur serta pengelolaan yang tepat. Kepadatan penduduk dapat diukur menggunakan perbandingan antara banyak orang dengan besar wilayah [3]. Bencana alam ialah peristiwa yang dapat mengancam dan mengganggu kehidupan manusia yang disebabkan oleh alam sehingga mengakibatkan korban jiwa kerusakan lainnya [4]. Banyaknya jumlah aliran air sungai pada suatu wilayah dapat menjadi faktor yang memengaruhi terjadinya banjir [5]. Banyaknya insiden yang terjadi di Indonesia termasuk di wilayah Jawa Barat yang mengakibatkan kerugian bagi warga setempat, baik itu harta hingga korban jiwa [6].

Permasalahan utama yang menjadi perhatian dalam penelitian ini adalah bagaimana persebaran kepadatan penduduk dan sebaran sungai di Jawa Barat dapat meningkatkan kerentanan terhadap bencana banjir, dan bagaimana pola tersebut dapat diidentifikasi serta dikelompokkan secara sistematis untuk mendukung upaya mitigasi.

Pada penelitian terdahulu juga menggunakan *Hierarchical Clustering* yang dilakukan oleh Ridzki Okta Pratikto [7] menunjukkan hasil informasi geografis yang sangat mengidentifikasi pada wilayah yang memiliki dampak banjir [7]. Dan juga penelitian yang telah dilakukan oleh Muhammad Ilham Mubarak [8] dalam menampilkan peta zonasi kepadatan penduduk untuk kasus kumulatif covid-19 di provinsi Sulawesi Selatan. Hal ini dapat membantu pemerintah dalam memahami pola penduduk, mengidentifikasi resiko bencana alam yang terjadi, merencanakan penanggulangan bencana, dan mengambil tindakan yang tepat. Oktari [9] mengusulkan bahwa strategi adaptasi kesehatan terkait pada perubahan iklim yang dapat membantu masyarakat yang rentan di Indonesia. Kebijakan semacam ini dapat membantu mengurangi dampak bencana di daerah dengan kepadatan penduduk yang tinggi, dimana resiko bencana lebih besar [9]. Kepadatan penduduk sering kali menyebabkan pembangunan permukiman yang tidak memperhatikan resapan air. Hal ini mengakibatkan pengurangan area yang dapat menyerap air hujan, sehingga meningkatkan risiko banjir. Penelitian ini secara spesifik menerapkan algoritma *Hierarchical Clustering*

untuk mengelompokkan wilayah-wilayah di provinsi Jawa Barat berdasarkan tingkat kepadatan penduduk dan jumlah aliran sungai. Dengan pendekatan ini, penelitian bertujuan untuk mengungkap pola kerentanan terhadap bencana secara spasial, sehingga pemerintah dapat merancang kebijakan mitigasi bencana yang lebih tepat sasaran. Dengan mengetahui pola kepadatan penduduk, dan sungai diharapkan dapat lebih memahami tingkat risiko bencana alam di berbagai wilayah Jawa Barat. Pada penelitian ini ada tiga fokus kelompok yaitu terpadat, sedang, dan terendah. Dalam penggunaan tiga kelompok itu akan mempermudah peneliti untuk memahami pengelompokan tersebut, dimana penelitian ini memiliki atribut kepadatan penduduk, dan sungai. Dengan demikian, informasi dari analisis *clustering* tentang pola kepadatan penduduk, dan sungai. Hal ini dapat digunakan sebagai dasar untuk merencanakan strategi penanggulangan bencana yang lebih efektif dan efisien di provinsi Jawa Barat.

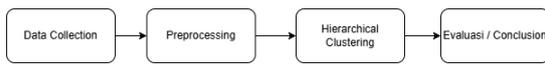
2. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik data mining yang menggunakan algoritma *Clustering Hierarki*. Tujuan dari penelitian ini adalah untuk mengelompokkan beberapa kabupaten dan kota di provinsi Jawa Barat berdasarkan tingkat kepadatan penduduknya. Klasifikasi ini didasarkan pada data sekunder resmi, seperti Badan Pusat Statistik (BPS). Penelitian ini mencakup pengumpulan dan pra-pemrosesan data, penerapan algoritma *clustering hierarchical*, dan evaluasi untuk membuat dendrogram yang menunjukkan hasil clustering. Untuk mengevaluasi seberapa baik hasil pengelompokan, *Davies-Bouldin Index* digunakan untuk melakukan evaluasi kualitas pengelompokan. Metode ini digunakan untuk menemukan pola-pola kepadatan yang dapat digunakan sebagai dasar untuk mempertimbangkan kebijakan pemerintah dan perencanaan wilayah.

2.1 Metode Analisis

Pada tahap ini ialah memilih metode data mining yang sesuai dengan tujuan penelitian. Metode yang digunakan adalah metode clustering (pengelompokan), Classification (klasifikasi), regression (regresi), association rule mining (aturan asosiasi), dan neural

networks (jaringan saraf). Metode yang digunakan adalah *clustering hierarchical*. Data mining merupakan suatu proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait dari berbagai database besar [10]. Data mining ialah subjek yang sangat interdisipliner, data mining dapat dipaparkan diberbagai teknologi secara nyata [11]. Metode yang digunakan dalam penulisan analisis data ini ialah metode *clustering hierarchical*. Tahapan analisis data yang dilakukan oleh penulis dalam penelitian ini dapat dilihat pada gambar 1.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Penelitian ini menggunakan data kepadatan penduduk di provinsi Jawa Barat dari tahun 2020 hingga 2022. Data tersebut diperoleh dari situs web resmi dinas kependudukan Jawa Barat <https://jabar.bps.go.id/indikator/12/245/1/kepadatan-penduduk-menurut-kabupaten-kota.html> [12]. Berdasarkan data dari dinas kependudukan ada 18 kabupaten dan 9 kota total data yaitu 27 data. Adapun atribut data yang digunakan pada penelitian ini ialah jumlah penduduk, dan sungai. Dari data kepadatan penduduk pada provinsi Jawa Barat dari tahun 2020 sampai tahun 2022. Data yang telah dikumpulkan akan diolah dengan metode pengelompokan menggunakan *clustering hierarchical* dan mendapatkan hasil pengelompokan yang tepat. Data ini akan dikelompokkan menjadi tiga kategori objek, yaitu "terpadat (C2)", "sedang (C1), dan "terendah (C0)".

Tabel 1. Data kepadatan penduduk Perkota/Kabupaten di BPS Jawa Barat

Kabupaten/Kota	Kepadatan Penduduk Menurut Kabupaten/Kota (jiwa/km)		
	2020	2021	2022
Bogor	2002	2025	1861
Sukabumi	657	666	674
Cianjur	645	653	700
Bandung	2050	2074	2136
Garut	841	847	847
Tasikmalaya	731	738	705
Ciamis	869	875	782
Kuningan	1051	1063	1003
Cirebon	2306	2327	2150
Majalengka	1084	1095	1004
Sumedang	759	764	745
Indramayu	899	907	902
Subang	842	849	750
Purwakarta	1208	1225	1036
Karawang	1476	1494	1309
Bekasi	2541	2578	2570
Bandung Barat	1370	1389	1439
Pangandaran	419	423	383
Kota Bogor	8802	8881	9550
Kota Sukabumi	7178	7271	7377
Kota Bandung	14577	14630	14776
Kota Cirebon	8921	9017	8646
Kota Bekasi	12311	12414	12159
Kota Depok	10267	10415	10622
Kota Cimahi	14474	14556	13557
Kota Tasikmalaya	4173	4218	3988
Kota Banjar	1771	1792	1576

Analisis perubahan yang lebih mendalam dan signifikan dapat dicapai dengan menggunakan data dari periode tiga tahun daripada data satu tahun. Hasil yang lebih stabil dan dapat diandalkan cenderung berasal dari data yang lebih luas. Kekuatan statistik untuk menemukan efek signifikan meningkat dengan jumlah sampel yang lebih besar, yang berarti bahwa hasil analisis lebih mungkin mencerminkan realitas daripada kebetulan semata. Penggunaan data selama tiga tahun dapat mengubah nilai atau *value* dalam analisis statistik. Dengan lebih banyak data, nilai *p* dapat menunjukkan signifikansi statistik pada tingkat yang sebelumnya tidak terlihat.

2.3 Preprocessing

Setelah data terkumpul, data akan menjalankan pemrosesan yang mencakup proses

pembersihan data, transformasi data, reduksi data, dan integrasi data untuk persiapan analisis data menggunakan algoritma data mining dengan tujuan untuk memperoleh data yang bermanfaat [13]. Pra-prosesing ini dilakukan dengan menggabungkan data dari sumber yang berbeda.

- **Pembersihan Data**
Pada tahap ini, data dibersihkan dari nilai yang hilang, outlier, dan inkonsistensi. Data yang tidak lengkap atau tidak akurat dapat menyebabkan hasil *clustering* yang tidak optimal.
- **Transformasi Data**
Transformasi data bertujuan untuk mengubah data ke dalam format yang lebih sesuai untuk algoritma *clustering*.
- **Reduksi Data**
Reduksi data bertujuan untuk mengurangi dimensi data tanpa kehilangan informasi penting.
- **Integrasi Data**
Pada tahap ini, data dari berbagai sumber digabungkan menjadi satu dataset yang koheren.

Pada kasus penelitian ini menggunakan pra-prosesing data dengan teknik *integrasi*. Data yang diintegrasikan adalah menggabungkan atribut banyaknya sungai, dan jumlah penduduk dari setiap wilayah pertahunnya.

2.4 Algoritma *Hierarchical Clustering*

Clustering merupakan salah satu teknik dalam data mining yang bertujuan untuk mengelompokkan objek-objek data berdasarkan kemiripan karakteristik tertentu, sehingga setiap kelompok (cluster) yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang jelas terhadap kelompok lainnya [14]. Salah satu pendekatan yang digunakan dalam penelitian ini adalah *hierarchical clustering*, yaitu metode pengelompokan data yang membentuk struktur hirarki atau pohon (dendrogram), di mana proses pengelompokan dilakukan secara bertingkat atau bertahap [15]. Metode ini tidak memerlukan penentuan jumlah cluster di awal, melainkan secara iteratif menggabungkan objek-objek data berdasarkan jarak terkecil hingga akhirnya membentuk satu cluster tunggal. Langkah awal dalam proses *hierarchical clustering* adalah perhitungan matriks jarak antar objek, yang pada penelitian

ini dilakukan menggunakan metode *Euclidean Distance*, untuk mengukur tingkat kemiripan antar data. Proses penggabungan dilakukan dengan menggabungkan pasangan cluster yang memiliki jarak terkecil secara berulang, hingga terbentuk struktur hirarki dari keseluruhan data. Algoritma ini secara bertahap menggabungkan N cluster awal menjadi satu cluster tunggal. Dalam kasus ini, *hierarchical clustering* akan menghasilkan 27 cluster, yang sesuai dengan jumlah objek atau data yang ada [15]. Pada tahap awal, setiap objek dianggap sebagai satu cluster tersendiri, sehingga untuk 27 objek data akan terbentuk 27 cluster awal yang kemudian digabungkan secara bertahap hingga terbentuk struktur yang utuh. Berbeda dengan algoritma clustering lainnya seperti k-means, metode *hierarchical clustering* tidak bergantung pada inisialisasi pusat cluster. Hal ini menjadi keunggulan utama, mengingat metode seperti k-means seringkali sensitif terhadap pemilihan nilai awal pusat cluster, yang dapat memengaruhi akurasi dan stabilitas hasil clustering secara keseluruhan. Hal ini menyebabkan hasil clusternya berupa solusi yang sifatnya local optimal [16].

$$d_{ik} = \sqrt{(X_i - Y_i)^2} \quad (1)$$

Dimana:

X_i = Pusat Awal (centroid)

Y_i = Data

ik = Jarak

Rumus ini digunakan dalam pengelompokan data (clustering) untuk menentukan pusat cluster terdekat, diterapkan dalam pengenalan pola atau klasifikasi untuk menghitung jarak antara vektor, dan bermanfaat dalam analisis data untuk mengukur kesamaan dan perbedaan antar data.

2.5 Evaluasi

Langkah terakhir adalah mengambil kesimpulan dari hasil analisis data mining. Dalam data mining juga evaluasi merupakan fase penting dalam mengevaluasi performa model yang dihasilkan. Evaluasi bertujuan untuk menilai keefektifan model atau penemuan yang diperoleh dari data mining. Proses ini melibatkan penerapan metrik evaluasi yang sesuai, seperti akurasi, presisi, recall, atau metrik lain yang relevan, tergantung pada jenis analisis yang dilakukan. Teknik evaluasi yang

dilakukan dalam penelitian ini menggunakan teknik matriks akurasi *Davies–Bouldin index*. Kesimpulan yang diambil harus dapat memberikan rekomendasi atau solusi untuk permasalahan atau tujuan penelitian.

3. HASIL DAN PEMBAHASAN

3.1 Pra-proses Data

Pada penelitian ini menggunakan pra-prosesing data dengan teknik *integrasi*. Data yang diintegrasikan adalah menggabungkan atribut banyaknya Sungai dan luas wilayah ke dalam data jumlah penduduk dari setiap wilayah. Data kepadatan penduduk untuk setiap kabupaten atau kota di provinsi Jawa Barat diperoleh dari <https://jabar.bps.go.id/indicator/12/245/1/kepadatan-penduduk-menurut-kabupaten-kota.html>. Pada saat ini, hanya data kepadatan tiga tahun yaitu 2020 sampai 2023 digunakan; tren multi-tahun, yang dapat membantu kita memahami lebih baik dinamika kependudukan, tidak dipertimbangkan. Data dibersihkan dan distandarisasi untuk menghilangkan suara dan memastikan format yang seragam. Penelitian ini menggabungkan data pertama yaitu kepadatan penduduk, dan juga banyaknya sungai serta luas wilayah tiap daerah kabupaten kota di provinsi Jawa Barat seperti pada Gambar 2.

Kabupaten/Kota	Kepadatan Penduduk Menurut Kabupaten/Kota (jiwa/km)			Jumlah sungai	Luas Wilayah/km
	2020	2021	2022		
Bogor	2002	2025	1861	70	118,50
Sukabumi	657	666	674	60	4.145,70
Cianjur	645	653	700	40	3.840,16
Bandung	2050	2074	2136	50	1.767,96
Garut	841	847	847	40	3.074,07
Tasikmalaya	731	738	705	30	2.551,19
Ciamis	869	875	782	30	1.414,71
Kuningan	1051	1063	1003	30	1.110,56
Cirebon	2306	2327	2150	20	37,36
Majalengka	1084	1095	1004	30	1.204,24
Sumedang	759	764	745	30	1.518,33
Indramayu	899	907	902	20	2.040,11
Subang	842	849	750	30	1.893,95
Purwakarta	1208	1225	1036	20	825,74
Karawang	1476	1494	1309	20	1.652,20
Bekasi	2541	2578	2570	20	1.224,88
Bandung Barat	1370	1389	1439	30	1.305,77
Pangandaran	419	423	383	10	1.010,00
Kota Bogor	8802	8881	9550	10	118,50
Kota Sukabumi	7178	7271	7377	10	48,25
Kota Bandung	14577	14630	14776	10	167,67
Kota Cirebon	8921	9017	8646	5	37,36
Kota Bekasi	12311	12414	12159	10	206,61
Kota Depok	10267	10415	10622	10	200,29
Kota Cimahi	14474	14556	13557	5	39,27
Kota Tasikmalaya	4173	4218	3988	10	171,61
Kota Banjar	1771	1792	1576	5	113,49

Gambar 2. Data Intregasi

3.2 Reduction Data

Adapun penelitian ini menggunakan data *reduction* dimana reduksi data adalah proses untuk menyederhanakan data dengan menghilangkan informasi yang tidak perlu.

Hasil clustering dapat diperjelas dan faktor pendorong kepadatan dapat diperjelas jika variabel tambahan seperti migrasi, urbanisasi, dan akses infrastruktur dimasukkan. Proses ini penting dalam data mining karena data yang disederhanakan akan lebih mudah dianalisis dan dipahami. Reduksi data dilakukan karena jenis data yang akan dicari ialah data kunjungan wisata ke objek wisata.

Tabel 2. Jumlah Data Rata-Rata Penduduk dan Sungai

No	Kabupaten/Kota	Penduduk	Sungai
1	Bogor	1962.666	94.25
2	Sukabumi	665.666	2102.85
3	Cianjur	666	1940.08
4	Bandung	2086.666	908.98
5	Garut	845	1557.035
6	Tasikmalaya	724.666	1290.595
7	Ciamis	842	722.355
8	Kuningan	1039	570.28
9	Cirebon	2261	28.68
10	Majalengka	1061	617.12
11	Sumedang	756	774.165
12	Indramayu	902.666	1030.055
13	Subang	813.666	961.975
14	Purwakarta	1156.666	422.87
15	Karawang	1426.666	836.1
16	Bekasi	2563	622.44
17	Bandung Barat	1399.333	667.885
18	Pangandaran	408.333	510
19	Kota Bogor	9077.666	64.25
20	Kota Sukabumi	7275.333	29.125
21	Kota Bandung	14661	88.835
22	Kota Cirebon	8861.333	21.18
23	Kota Bekasi	12294.666	108.305
24	Kota Depok	10434.666	105.145
25	Kota Cimahi	14195.666	22.135
26	Kota Tasikmalaya	4126.333	90.805
27	Kota Banjar	1713	59.245

3.3 Metode Data Mining Penentuan Pusat Awal Clustering

Sebelum memproses data ada beberapa tahapan data mining yang dilakukan yaitu:

- Pengumpulan data : mengambil data kependudukan dari BPS Jawa barat
- Processing data : Membersihkan data, mengisi data yang kosong, menyamakan format, normalisasi nilai.
- Pemilihan atribut : Menggunakan dua atribut utama yaitu kepadatan penduduk dan jumlah sungai.

- Pemilihan algoritma clustering : Menggunakan hierarchical clustering karena tidak memerlukan jumlah cluster di awal dan cocok untuk analisis struktur.
- Evaluasi hasil clustering : Menggunakan *Davies-Bouldin Index* untuk menilai kualitas cluster yang terbentuk.
- Interpretasi hasil dan pengelompokan wilayah : Menjelaskan kenapa kota tertentu masuk cluster terpadat, sedang, terendah dan bagaimana itu berdampak pada risiko banjir.

Proses pengelompokan data dilakukan dengan mencari nilai rata-rata untuk setiap kabupaten/kota terlebih dahulu.

$$R1 = 1962.666 + 94.25 / 2 = 1028.458$$

$$R2 = 665.666 + 2102.85 / 2 = 1384.258$$

$$R3 = 666 + 1940.08 / 2 = 1303.04$$

$$R4 = 2086.666 + 908.08 / 2 = 1497.823$$

$$R5 = 845 + 1557.035 / 2 = 1201.018$$

Perhitungan untuk mencari nilai rata-rata berlanjut hingga R27. Selama proses pengelompokan, data yang diperoleh dihitung rata-rata terlebih dahulu berdasarkan jumlah penduduk atau sungai. Data tersebut kemudian dijumlahkan dan ditentukan nilai rata-ratanya:

Tabel 3. Jumlah Rata-Rata

No	Kabupaten/Kota	Rata-Rata
1	Bogor	1028.458
2	Sukabumi	1384.258
3	Cianjur	1303.04
4	Bandung	1497.823
5	Garut	1201.018
6	Tasikmalaya	1007.631
7	Ciamis	782.1775
8	Kuningan	804.64
9	Cirebon	1144.84
10	Majalengka	839.06
11	Sumedang	765.0825
12	Indramayu	966.3605
13	Subang	887.8205
14	Purwakarta	789.768
15	Karawang	1131.383
16	Bekasi	1592.72
17	Bandung Barat	1033.609
18	Pangandaran	459.1665
19	Kota Bogor	4570.958
20	Kota Sukabumi	3652.229
21	Kota Bandung	7374.918

22	Kota Cirebon	4441.257
23	Kota Bekasi	6201.486
24	Kota Depok	5269.906
25	Kota Cimahi	7108.901
26	Kota Tasikmalaya	2108.569
27	Kota Banjar	886.1225

Setelah data diperoleh nilai rata – ratanya, langkah selanjutnya ialah menentukan pusat awal atau centroid awalnya. Hal ini ditentukan secara acak dengan menggunakan nilai cluster nol pada baris 18 (Pangandaran) dan nilai cluster satu pada baris 9 (Cirebon) dan nilai cluster dua berada pada baris 21 (Kota Bandung).

Tabel 4. Centroid Data Awal

C0	18	Pagandaran	459.16
C1	9	Cirebon	1144.84
C2	21	Kota Bandung	7374.918

3.4 Algoritma Hierarchical Clustering

Tahap selanjutnya pada penelitian ini ialah penggunaan metode hierarchical clustering dimana teknik clustering ini membentuk hirarki sehingga membentuk struktur pohon. Algoritma clustering hierarkis dipilih karena mampu menunjukkan struktur hierarki antar wilayah yang didasarkan pada kemiripan kepadatan penduduk. *Hierarchical clustering* memiliki kompleksitas komputasi yang lebih tinggi daripada metode non-hierarki seperti K-Means. Dengan demikian proses pengelompokannya dilakukan secara bertingkat atau bertahap. Sebelum memulai proses clustering, pertamanya kita perlu menghitung matriks jarak antara setiap pasangan data menggunakan rumus Euclidean Distance.

$$d_{ik} = \sqrt{(X_i - Y_i)^2} \quad (2)$$

Dimana:

X_i = Pusat Awal (centroid)

Y_i = Data

ik = Jarak

Kemudian dapat nilai matriks sebagai berikut :

Jarak data 1 ke pusat cluster

$$C0 = \sqrt{(459.16665 - 1028.458)^2} = 324092.81$$

$$C1 = \sqrt{(1144.84 - 1028.458)^2} = 13544.77$$

$$C2 = \sqrt{(7374.918 - 1028.458)^2} = 40277554.53$$

Jarak data 2 ke pusat cluster

$$C0 = \sqrt{(459.16665 - 1384.258)^2} = 855794.28$$

$$C1 = \sqrt{(1144.84 - 1384.258)^2} = 57320.97$$

$$C2 = \sqrt{(7374.918 - 1384.258)^2} = 35888007.24$$

Jarak data 3 ke pusat cluster

$$C0 = \sqrt{(459.16665 - 1303.04)^2} = 712122.48$$

$$C1 = \sqrt{(1144.84 - 1303.04)^2} = 25027.24$$

$$C2 = \sqrt{(7374.918 - 1303.04)^2} = 36867702.45$$

Jarak data 4 ke pusat cluster

$$C0 = \sqrt{(459.16665 - 1497.823)^2} = 1078807.32$$

$$C1 = \sqrt{(1144.84 - 1497.823)^2} = 124596.99$$

$$C2 = \sqrt{(7374.918 - 1497.823)^2} = 34540245.64$$

Jarak data 5 ke pusat cluster

$$C0 = \sqrt{(459.16665 - 1201.018)^2} = 550342.91$$

$$C1 = \sqrt{(1144.84 - 1201.018)^2} = 3155.91$$

$$C2 = \sqrt{(7374.918 - 1201.018)^2} = 38117041.21$$

Dan seterusnya sampai dengan data ke-27 terhadap pusat awal cluster hingga mendapatkan matriks jarak. Berikut adalah gambar 3 yang berisi hasil perhitungan jarak antara data dengan titik pusat pada tabel literasi satu :

NO	Kabupaten/Kota	Rata - Rata	Literasi 1			Jarak Terpendek
			C0	C1	C2	
1	Bogor	1028.458	324092.81	13544.77	40277554.53	13544.77
2	Sukabumi	1384.258	855794.28	57320.97	35888007.24	57320.98
3	Cianjur	1303.04	712122.48	25027.24	36867702.45	25027.24
4	Bandung	1497.823	1078807.32	124596.99	34540245.64	124597.00
5	Garut	1201.018	550342.91	3155.91	38117041.21	3155.91
6	Tasikmalaya	1007.631	300813.3078	18826.31	40542343.74	18826.31
7	Ciamis	782.1775	104336.11	131524.09	43464227.3	104336.11
8	Kuningan	804.64	119351.93	115736.04	43168553	115736.04
9	Cirebon	1144.84	470148.15	0.00	38813871.89	0.00
10	Majalengka	839.06	144319.07	93501.41	42717439.8	93501.41
11	Sumedang	765.0825	93584.59	144215.76	43689925.34	93584.60
12	Indramayu	966.3605	257245.75	31854.93	41069609.23	31854.93
13	Subang	887.8205	183744.25	66059.02	42082433.97	66059.02
14	Purwakarta	789.768	109297.35	126076.12	43364200.52	109297.35
15	Karawang	1131.383	451875.02	181.09	38981729.3	181.09
16	Bekasi	1592.72	1284943.54	200596.49	33433813.71	200596.49
17	Bandung Barat	1033.609	329984.18	12372.34	40212199.83	12372.34
18	Pangandaran	459.1665	0.00	470148.14	47827618.81	0.00
19	Kota Bogor	4570.958	16906829.34	11738284.55	7862191.68	7862191.68
20	Kota Sukabumi	3652.229	10195648.13	6286999.60	13858413.39	6286999.60
21	Kota Bandung	7374.918	47827611.89	38813865.66	0	0.00
22	Kota Cirebon	4441.257	15857044.75	10866365.04	8606366.86	8606366.86
23	Kota Bekasi	6201.486	32974227.50	25569663.71	1376942.66	1376942.66
24	Kota Depok	5269.906	23143214.54	17016169.5	4431075.52	4431075.52
25	Kota Cimahi	7108.901	44218962.27	35570017.65	70765.04	70765.04
26	Kota Tasikmalaya	2108.569	2720528.60	928773.59	27734431.79	928773.59
27	Kota Banjar	886.1225	182291.43	66934.74	42104467.04	66934.74

Gambar 3. Literasi satu

Nilai SSE yaitu :

$$SSE = \sum_{i=1}^k \sum_{p \in C} d(p, m_i)^2 \quad (3)$$

$$13544.772 + 57320.982 + 25027.242 + 124597.002 + 3155.912 + 18826.312 + 104336.112 + 115736.042 + 0.002 + 93501.412 + 93584.602 + 31854.932 + 66059.022 + 109297.352 + 181.092 + 200596.492 + 12372.342 + 0.002 + 7862191.682 + 6286999.602 + 0.002 + 8606366.862 + 1376942.662 + 4431075.522 + 70765.042 + 928773.592 + 66934.742 = 30700041.282$$

Pada *Agglomerative* cara kerjanya dimulai dari mengelompokkan objek objek individual yang mana pada awalnya jumlah cluster samadengan banyaknya jumlah objek. Lalu objek objek yang memiliki kemiripan atau kedekatan dikelompokkan membentuk kelompok baru berdasarkan hasil perhitungan jarak dan parameter kedekatan yang digunakan. Parameter yang digunakan ialah *distance matrix* yang dimana rumus yang digunakan untuk mengukur jarak antara dua titik data sehingga jenis *distance matrix* yang digunakan dalam penelitian ini adalah *euclidean distance*. Selanjutnya dilakukan perhitungan jarak antar cluster dan menggabungkan cluster yang memiliki jarak terdekat. Begitu seterusnya hingga seluruh objek membentuk satu cluster.

3.5 Pengelompokan Data

Untuk menentukan kelompok data, jarak antara data dan pusat cluster dihitung terlebih dahulu. Data yang memiliki jarak paling dekat dengan pusat cluster akan dikelompokkan ke dalam cluster yang sama dengan pusat cluster tersebut. Tabel 5 menunjukkan pengelompokan data, dengan nilai literasi satu berarti data tersebut berada dalam kelompok.

Tabel 5. Hasil Cluster Literasi 1

No	Kabupaten/Kota	Literasi 1		
		C0	C1	C2
1	Bogor	1		
2	Sukabumi	1		
3	Cianjur	1		
4	Bandung	1		
5	Garut	1		
6	Tasikmalaya	1		
7	Ciamis	1		
8	Kuningan	1		
9	Cirebon	1		
10	Majalengka	1		
11	Sumedang	1		
12	Indramayu	1		
13	Subang	1		
14	Purwakarta	1		
15	Karawang	1		
16	Bekasi	1		
17	Bandung Barat	1		
18	Pangandaran	1		
19	Kota Bogor		1	
20	Kota Sukabumi		1	
21	Kota Bandung			1

22	Kota Cirebon	1
23	Kota Bekasi	
24	Kota Depok	
25	Kota Cimahi	
	Kota	1
26	Tasikmalaya	
27	Kota Banjar	1

Setelah diketahui data anggota dari setiap cluster, pusat cluster baru dihitung berdasarkan data tersebut. Rumus yang digunakan untuk menghitung pusat cluster baru adalah sebagai berikut:

$$C0 = \frac{13544.77 + 57320.98 + 25027.24 + 124597.00 + 3155.91 + 18826.31 + 104336.11 + 115736.04 + 0.00 + 93501.41 + 93584.60 + 31854.93 + 66059.02 + 109297.35 + 181.09 + 200596.49 + 12372.34 + 0.00 + 928773.59 + 66934.74}{20} = 103.284.99$$

$$C1 = \frac{7862191.68 + 6286999.60 + 8606366.86}{3} = 7.585.186.05$$

$$C2 = \frac{47827611.89 + 1376942.66 + 4431075.52 + 70765.04}{4} = 13.426.598.78$$

Setelah titik tengah cluster baru ditentukan, langkah selanjutnya adalah menghitung jarak antara setiap data jumlah wisatawan berdasarkan kabupaten/kota dengan masing-masing titik tengah cluster. Hasil perhitungan jarak ini akan disimpan dalam matriks jarak sebagai berikut :

Jarak data 1 ke pusat cluster

$$C0 = \sqrt{(103284.99 - 1028.458)^2} = 102256.532$$

$$C1 = \sqrt{(7585186.05 - 1028.458)^2} = 7584157.592$$

$$C2 = \sqrt{(13426598.78 - 1028.458)^2} = 13425570.32$$

Jarak data 2 ke pusat cluster

$$C0 = \sqrt{(103284.99 - 1384.258)^2} = 101900.732$$

$$C1 = \sqrt{(7585186.05 - 1384.258)^2} = 7583801.792$$

$$C2 = \sqrt{(13426598.78 - 1384.258)^2} = 13425214.52$$

Jarak data 3 ke pusat cluster

$$C0 = \sqrt{(103284.99 - 1303.04)^2} = 101981.95$$

$$C1 = \sqrt{(7585186.05 - 1303.04)^2} = 7583883.01$$

$$C2 = \sqrt{(13426598.78 - 1303.04)^2} = 13425295.74$$

Jarak data 4 ke pusat cluster

$$C0 = \sqrt{(103284.99 - 1497.823)^2} = 101787.167$$

$$C1 = \sqrt{(7585186.05 - 1497.823)^2} = 7583688.227$$

$$C2 = \sqrt{(13426598.78 - 1497.823)^2} = 13425100.96$$

Jarak data 5 ke pusat cluster

$$C0 = \sqrt{(103284.99 - 1201.018)^2} = 102083.972$$

$$C1 = \sqrt{(7585186.05 - 1201.018)^2} = 102083.972$$

$$C2 = \sqrt{(13426598.78 - 1201.018)^2} = 13425397.76$$

Dan seterusnya sampai dengan data ke-27 terhadap pusat awal cluster hingga mendapatkan matriks jarak. Gambar 4 menunjukkan hasil

perhitungan jarak antara data dengan titik pusat pada literasi dua:

NO	Kabupaten/Kota	Rata - Rata	Literasi 2			Jarak Terpendek
			C0	C1	C2	
1	Bogor	1028.458	102256.532	7584157.592	13425570.32	102256.532
2	Sukabumi	1384.258	101900.732	7583801.792	13425214.52	101900.732
3	Cianjur	1303.04	101981.95	7583883.01	13425295.74	101981.950
4	Bandung	1497.823	101787.167	7583688.227	13425100.96	101787.167
5	Garut	1201.018	102083.972	7583985.032	13425397.76	102083.972
6	Tasikmalaya	1007.631	102277.359	7584178.419	13425991.15	102277.359
7	Ciamis	782.1775	102502.812	7584403.873	13425816.6	102502.812
8	Kuningan	804.64	102480.35	7584381.41	13425794.14	102480.350
9	Cirebon	1144.84	102140.15	7584041.21	13425453.94	102140.150
10	Majalengka	839.06	102445.93	7584346.99	13425759.72	102445.930
11	Sumedang	765.0825	102528.908	7584429.968	13425842.7	102528.908
12	Indramayu	966.3605	102318.63	7584219.689	13425632.42	102318.630
13	Subang	887.8205	102397.17	7584298.229	13425710.96	102397.170
14	Purwakarta	789.768	102495.222	7584396.282	13425809.01	102495.222
15	Karawang	1131.383	102153.607	7584054.667	13425467.4	102153.607
16	Bekasi	1592.72	101692.27	7583593.33	13425006.06	101692.270
17	Bandung Barat	1033.609	102251.381	7584152.441	13425565.17	102251.381
18	Pangandaran	459.1665	102825.824	7584726.883	13426139.61	102825.824
19	Kota Bogor	4570.958	7580615.09	98714.03	13422027.82	98714.032
20	Kota Sukabumi	3652.229	7581533.82	99632.76	13422946.55	99632.761
21	Kota Bandung	7374.918	7577811.13	13419223.86	95910.07	95910.072
22	Kota Cirebon	4441.257	7580744.79	98843.73	13422157.52	98843.733
23	Kota Bekasi	6201.486	7578984.56	13420397.29	97083.50	97083.504
24	Kota Depok	5269.906	7579916.14	13421328.87	98015.08	98015.084
25	Kota Cimahi	7108.901	7578077.15	13419489.88	96176.09	96176.089
	Kota					
26	Tasikmalaya	2108.569	101176.42	7583077.48	13424490.21	101176.421
27	Kota Banjar	886.1225	102398.87	7584299.93	13425712.66	102398.868

Gambar 4. Literasi 2

Berikut adalah hasil perhitungan yang dilakukan pada data cluster literasi dua.

Tabel 6. Hasil Cluster Literasi Dua

No	Kabupaten/Kota	Literasi 2		
		C0	C1	C2
1	Bogor	1		
2	Sukabumi	1		
3	Cianjur	1		
4	Bandung	1		
5	Garut	1		
6	Tasikmalaya	1		
7	Ciamis	1		
8	Kuningan	1		
9	Cirebon	1		
10	Majalengka	1		
11	Sumedang	1		
12	Indramayu	1		
13	Subang	1		
14	Purwakarta	1		
15	Karawang	1		
16	Bekasi	1		
17	Bandung Barat	1		
18	Pangandaran	1		
19	Kota Bogor		1	
20	Kota Sukabumi		1	
21	Kota Bandung			1
22	Kota Cirebon		1	
23	Kota Bekasi			1
24	Kota Depok			1
25	Kota Cimahi			1
	Kota			
26	Tasikmalaya	1		
27	Kota Banjar	1		

Nilai SSE yaitu:

$$SSE = \sum_{i=1}^k \sum_{p \in C} d(p, m_i)^2 \quad (4)$$

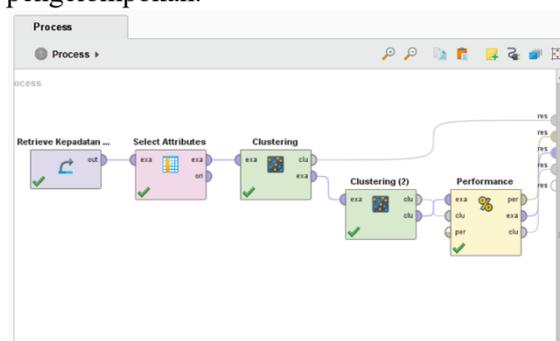
102256.532 + 101900.732 + 101981.950 +
 101787.167 + 102083.972 + 102277.359 +
 102502.812 + 102480.350 + 102140.150 +
 102445.930 + 102528.908 + 102318.630 +
 102397.170 + 102495.222 + 102153.607 +
 101692.270 + 102251.381 + 102825.824 +
 98714.032 + 99632.761 + 95910.072 +
 98843.733 + 97083.504 + 98015.084 +
 96176.089 + 101176.421 + 102398.868 =
 2728470.530

Setelah mendapatkan hasil literasi dua, hasil tersebut dibandingkan dengan hasil literasi satu. Proses pengelompokan data akan terus berlanjut selama titik pusat cluster masih berubah. Jika titik pusat cluster tidak lagi berubah, maka proses akan berhenti. Berdasarkan perhitungan manual pada data jumlah penduduk provinsi Jawa Barat berdasarkan kabupaten/kota, diperoleh hasil akhir literasi 2 yang sama, yaitu C0=20, C1=3, dan C2=4 dengan nilai SSE 2.728.470.530. Proses pengelompokan data berhenti karena hasil pengelompokan data pada literasi satu dan dua sama. Berdasarkan hasil pengelompokan data tersebut, disimpulkan bahwa Kabupaten/Kota dengan jumlah penduduk terendah: 20 Kabupaten/Kota, yaitu Bogor Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Tasikmalaya, dan Kota Banjar. Kabupaten/Kota dengan jumlah penduduk sedang: tiga Kabupaten/Kota, yaitu Kota Bogor, Sukabumi, dan Kota Cirebon. Kabupaten/Kota dengan Jumlah penduduk terbanyak: empat Kabupaten/Kota, yaitu Kota Bandung, Kota Bekasi, Kota Depok, dan Kota Cimahi.

3.6 Rapid Miner

Pada penelitian ini Rapid Miner digunakan sebagai alat bantu utama dalam penerapan teknik data mining, khususnya untuk melakukan proses clustering menggunakan metode *Hierarchical Clustering*. Rapid Miner merupakan sebuah platform analisis data berbasis visual yang mendukung berbagai teknik *machine learning*, termasuk klasifikasi, regresi, clustering, asosiasi, dan analisis teks.

Dalam penggunaan Rapid Miner lakukan *import* data yang ingin dianalisis, selanjutnya tambahkan *Select Attributes* untuk menentukan atribut yang akan dipakai. Setelah menentukan atributnya kemudian seret operator *Agglomerative Clustering* ke dalam lembar kerja. Lalu tambahkan operator *Flatten Clustering* untuk menentukan jumlah Cluster. Hubungkan data : hubungkan operator "retrieve data" dengan *Select Attributes* lalu operator *Agglomerative Clustering* lalu *Flatten Clustering* untuk proses pengelompokan. Pilih Jumlah Cluster: Tetapkan jumlah cluster yang sesuai dengan karakteristik data. Terakhir Evaluasi Performa: Tambahkan *Cluster Distance Performance* untuk melihat hasil pengelompokan.



Gambar 5. Metode Hierarchical Clustering pada RapiMiner

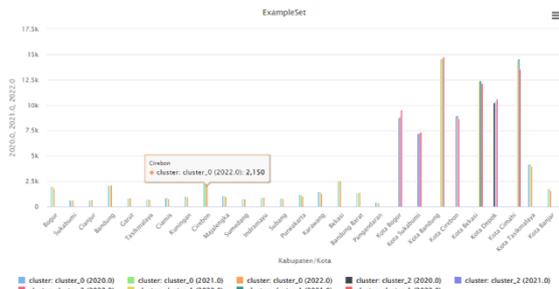
Dimana Hasil dari pengelompokan ini dapat dilihat pada gambar 6.

Cluster Model

```
Cluster 0: 20 items
Cluster 1: 3 items
Cluster 2: 4 items
Total number of items: 27
```

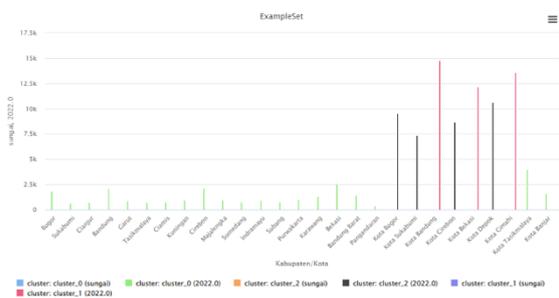
Gambar 6. Hasil Cluster

Dimana menunjukkan bahwa 20 kabupaten/kota dengan jumlah kepadatan penduduk terendah (Cluster 0). Sementara itu, 3 kabupaten/kota dengan jumlah kepadatan penduduk sedang (Cluster 1). Dan, 4 Kabupaten/kota dengan jumlah kepadatan penduduk yang terpadat (Cluster2).



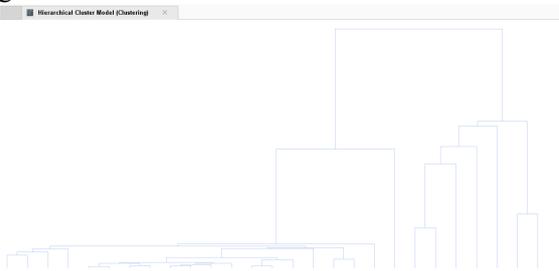
Gambar 7. Grafik Hasil dari Cluster Kepadatan Penduduk

Sedangkan untuk hasil cluster sungai dimana 20 kabupaten/kota dengan minim terjadinya dampak banjir, dan 4 kabupaten/kota dengan jarang terjadinya dampak banjir, dan 3 kabupaten/kota dengan sering terjadi dampak banjir seperti pada gambar 8.



Gambar 8. Hasil Cluster untuk daerah Sungai

Dengan N adalah jumlah data atau bisa dibilang jumlah awal cluster pada *Agglomerative Hierarchical Clustering* adalah sebanyak jumlah objek/data yaitu 53 cluster. Visualisasi hasil clustering dapat dilihat pada dendrogram gambar 9.



Gambar 9. Visualisasi Cluster dengan Dendrogram

3.7 Evaluasi

Evaluasi performa terakhir dalam penelitian ini dilakukan menggunakan teknik matriks *Davies–Bouldin Index*. *Davies-Bouldin Index* adalah metrik evaluasi yang digunakan untuk menilai kualitas hasil klasterisasi.

Davies Bouldin

Davies Bouldin: 0.112

Gambar 10. Evaluasi Performa *Davies–Bouldin Index*

Dimana ini adalah nilai *Davies-Bouldin index*, yang merupakan ukuran kualitas pengelompokan. Semakin kecil nilainya, semakin baik pengelompokan. Dalam gambar ini, nilainya adalah 0,112.

4. KESIMPULAN

Pada penelitian ini berhasil menerapkan algoritma *hierachical clustering* untuk mengidentifikasi kepadatan penduduk dan banyaknya sungai yang memiliki potensi bencana alam di provinsi Jawa Barat. Berdasarkan pengujian dan penerapan dari algoritma *hierachical clustering* untuk mengelompokkan kepadatan penduduk dan potensi banjir di provinsi Jawa Barat yang diuji dengan metode evaluasi *Davies-bouldin index* maka disimpulkan bahwa algoritma hierarchical clustering berhasil diterapkan, dimana untuk kepadatan penduduk 20 kabupaten/kota dengan jumlah kepadatan penduduk terendah (Cluster 0). Sementara itu, 3 kabupaten/kota dengan jumlah kepadatan penduduk Sedang (Cluster 1) dan, 4 Kabupaten/kota dengan jumlah kepadatan penduduk yang terpadat (Cluster 2). Dan untuk potensi banjir didapat pada 20 kabupaten/kota dengan minim terjadinya dampak banjir, dan 4 kabupaten/kota dengan jarang terjadinya dampak banjir, dan 3 kabupaten/kota dengan sering terjadi dampak banjir. Evaluasi hasil clustering dilakukan dengan metode akurasi *Davies-bouldin index* dengan nilai 0,112. Dengan begitu, potensi terjadi bencana alam di provinsi Jawa Barat terdapat pada cluster 3 yaitu kota Bandung, kota Bekasi, dan juga kota Cimahi sehingga sangat perlu untuk acuan diperhatikan oleh pemerintah Jawa Barat. Dengan demikian, informasi dari analisis clustering tentang pola kepadatan penduduk, dan sungai. Hasil ini memiliki konsekuensi signifikan bagi pemerintah provinsi Jawa Barat dan lembaga yang berhubungan dengannya. Dalam strategi penanggulangan bencana, kota-kota yang

terletak di wilayah dengan kepadatan tinggi dan rawan banjir harus menjadi prioritas utama. Ini termasuk membangun tata ruang wilayah berbasis risiko, memperkuat sistem drainase dan memproyeksikan daerah rawan banjir, dan meningkatkan kesadaran masyarakat terhadap bencana. Hal ini dapat digunakan sebagai dasar untuk merencanakan strategi penanggulangan bencana yang lebih efektif dan efisien di provinsi Jawa Barat. Agar hasil yang diperoleh lebih akurat dan bermanfaat untuk pengambilan kebijakan jangka panjang, peneliti harus menambahkan variabel curah hujan, tata guna lahan, dan tingkat urbanisasi.

DAFTAR PUSTAKA

- [1] C. Christiani, P. Tedjo, and B. Martono, "ANALISIS DAMPAK KEPADATAN PENDUDUK TERHADAP KUALITAS HIDUP MASYARAKAT PROVINSI JAWA TENGAH 1," 2014.
- [2] S. P. Nur Azizah, L. S. Pratiwi, I. Amaliah, and F. Fitriyana, "Sanitasi Dan Kepadatan Penduduk Sebagai Dinamika Kemiskinan Kota Studi Kasus Provinsi Jawa Barat", *JNAJPM*, vol. 7, no. 1, pp. 55–70, Apr. 2022..
- [3] T. A. Azis, A. S. Saputra, S. N. Azkiya, and C. Windiati, "Pengaruh Kepadatan Penduduk Terhadap Produk Domestik Regional Bruto (PDRB) Per Kapita Provinsi Jawa Barat Tahun 2022," vol. 3, no. 2, pp. 119–127, 2024, doi: 10.30640/inisiatif.v3i2.2265.
- [4] S. Hardiyanto, D. Pulungan, U. Muhammadiyah, and T. Selatan, "Komunikasi Efektif Sebagai Upaya Penanggulangan Bencana Alam di Kota Padangsidempuan," 2019.
- [5] N. Sulaksana, P. P. R. Rendra, and M. Sulastri, "SOSIALISASI MITIGASI BENCANA LONGSOR DAN BANJIR SECARA VIRTUAL DI MASA PANDEMI COVID-19," *Kumawula: Jurnal Pengabdian Kepada Masyarakat*, vol. 4, no. 3, p. 465, Dec. 2021, doi: 10.24198/kumawula.v4i3.35516.
- [6] P. Agina Widyaswara Suwaryo *et al.*, "Faktor-Faktor Yang Mempengaruhi Tingkat Pengetahuan Masyarakat dalam Mitigasi Bencana Alam Tanah Longsor," 2017.
- [7] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, R. O. Pratikto, and N. Damastuti, "Terakreditasi SINTA Peringkat 4 Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," 2018.
- [8] M. Ilham Mubarak *et al.*, "PEMETAAN ZONA RESIKO PENULARAN COVID-19 DI SULAWESI SELATAN MENGGUNAKAN PLOT DENDROGRAM HIERARCHICAL CLUSTERING (Mapping Of The Risk Zone Of Transmission Of Covid-19 In South Sulawesi Using A Hierarchical Clustering Dendrogram Plot)," 2021.
- [9] R. S. Oktari *et al.*, "Indonesia's Climate-Related Disasters and Health Adaptation Policy in the Build-Up to COP26 and Beyond," Jan. 01, 2022, *MDPI*. doi: 10.3390/su14021006.
- [10] S. Widaningsih, "PERBANDINGAN METODE DATA MINING UNTUK PREDIKSI NILAI DAN WAKTU KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DENGAN ALGORITMA C4,5, NAÏVE BAYES, KNN DAN SVM," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 16–25, Apr. 2019, doi: 10.36787/jti.v13i1.78.
- [11] J. Jaya Purnama and S. Rahayu, "KLASIFIKASI KONSUMSI ENERGI INDUSTRI BAJA MENGGUNAKAN TEKNIK DATA MINING," 2022. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- [12] BPS Provinsi Jawa Barat, "<https://jabar.bps.go.id/indicator/12/2>

- 45/1/kepadatan-penduduk-menurut-kabupaten-kota.html”.
- [13] B. S. Purnomo and P. T. Prasetyaningrum, “PENERAPAN DATA MINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN DI KOTA YOGYAKARTA MENGGUNAKAN METODE K-MEANS,” 2021.
- [14] M. Nishom and M. Y. Fathoni, “Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritma K-Means Clustering,” vol. 03, no. 02, 2018.
- [15] J. Homepage, K. Pratama Simanjuntak, and U. Khaira, “MALCOM: Indonesian Journal of Machine Learning and Computer Science Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm Pengelompokkan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering,” vol. 1, pp. 7–16, 2021.
- [16] T. Alfina, B. Santosa, and R. A. Barakbah, “Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS),” Sep. 2012.