

Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC)

Hot Spot Clustering Using Agglomerative Hierarchical Clustering (AHC) Algorithm

Rajnaparmitha Kusumastuti¹, Elfandry Bayunanda², Anggi Muhammad Rifa'i³,
Muhamad Ryandy Ghonim Asgar⁴, Fahma Inti Ilmawati⁵, Kusrini⁶

^{1,2,3,4,5}, Magister Teknik Informatika, Universitas Amikom, Yogyakarta

e-mail: ¹rajna@students.amikom.ac.id, ²bayunanda555@students.amikom.ac.id,
³anggi.rifai@students.amikom.ac.id, ⁴muhammadryandy@students.amikom.ac.id,
⁵fahma.inti.ilmawati@students.amikom.ac.id, ⁶kusrini@amikom.ac.id

Abstrak

Kebakaran hutan di Indonesia setiap tahunnya masih sering terjadi. Dalam menanggulangi kebakaran hutan sendiri, para peneliti belakangan ini semakin fokus untuk melakukan pengembangan sistem yang mampu melakukan prediksi kebakaran hutan. Selain melakukan prediksi mengenai kebakaran hutan, antisipasi yang dapat dilakukan untuk membantu menangani pencegahan kebakaran hutan salah satunya adalah melakukan pengelompokan terhadap wilayah yang memiliki potensi kebakaran. Pada penelitian ini melakukan Clusterisasi titik panas (hotspot) untuk membagi wilayah yang berpotensi untuk terbakar. Pengelompokan wilayah dilakukan berdasarkan cluster kebakaran rendah, sedang, dan tinggi. Clusterisasi wilayah dilakukan menggunakan algoritma Agglomerative Hierarchical Clustering (AHC). Data yang digunakan pada perhitungan ini menggunakan wilayah Kalimantan Barat dengan menggunakan variable longitude, latitude, frp, confidence, dan curah hujan untuk menentukan clustering wilayah kebakaran. Tujuan dalam penelitian ini adalah untuk clusterisasi titik panas kebakaran hutan berdasarkan wilayah yang berpotensi terjadi kebakaran hutan sehingga memiliki peluang untuk segera ditindaklanjuti. Dari hasil pengujian diperoleh pembentukan 2 cluster yang dimana menunjukkan terbentuk 2 pengelompokan titik panas (hotspot) pada kelas sedang dan tinggi dengan nilai evaluasi silhouette coefficient 0,771.

Kata kunci—Clustering, titik panas, Agglomerative Hierarchical Clustering

Abstract

Forest fires in Indonesia are still common every year. In tackling forest fires, researchers have recently been increasingly focused on developing systems that can predict forest fires. In addition to making predictions about forest fires, anticipation that can be done to help prevent forest fires, one of which is to group areas that have the potential for fires. In this study, hotspots are clustered to divide areas that have the potential to burn. Regional grouping is based on low, medium, and high fire clusters. Regional clustering was carried out using the Agglomerative Hierarchical Clustering (AHC) algorithm. The data used in this calculation uses the West Kalimantan area using the variables longitude, latitude, frp, confidence, and rainfall to determine the clustering of fire areas. The purpose of this study is to cluster forest fire hotspots based on areas that have the potential for forest fires to occur so that they have the opportunity to be followed up immediately. From the test results obtained the formation of 2 clusters which shows that 2 hotspots are formed in the medium and high class with a silhouette coefficient evaluation value of 0.771.

Keywords—Clustering, hotspots, Agglomerative Hierarchical Clustering

1 PENDAHULUAN

Indonesia terletak pada garis katulistiwa yang menjadikan negara ini memiliki iklim tropis yang Sebagian wilayahnya tergolong beriklim panas. Indonesia merupakan salah satu negara yang masih memiliki hutan lebat di dunia sekaligus masih meliki tingkat kebakaran hutan yang tinggi. Kebakaran adalah salah satu bencana yang kerap kali melanda Indonesia, khususnya daerah dengan hutan dan lahan yang luas seperti wilayah Kalimantan. Kebakaran hutan dan lahan menyebabkan kerusakan terhadap vegetasi hutan, kerusakan lingkungan, dan hal merugikan lainnya seperti pencemaran udara karena asap.

Berdasarkan hasil laporan dari Badan Nasional Penanggulangan Bencana (BNPB) pada tahun 2019, tercatat total kawasan yang terkena dampaknya mencapai 942.484 hektar. Jumlah ini mengalami kenaikan jika dibandingkan dengan tahun 2018 yang mencapai 529.266 hektar, dan 2017 mencapai 165.483,92 hektar. Kalimantan Tengah (161.298 hektar), Kalimantan Barat (131.654 hektar), NTT (120.143 hektar), Kalimantan Selatan (115.317 hektar), dan Sumatera Selatan (92.635 hektar). Dengan sebaran *hotspot* yang meliputi Kalimantan Tengah (39.634 titik), Kalimantan Barat (25.461 titik), Sumatera Selatan (22.457 titik), Jambi (12.491 titik), dan Riau (12.270 titik) [1]. Kebakaran hutan memiliki dampak yang sangat buruk bagi kesehatan, ekonomi, dan bidang lainnya. Kebakaran hutan masih terjadi setiap tahunnya di Indonesia, khususnya di daerah Kalimantan Barat. Tingginya tingkat kebakaran hutan yang terjadi mempengaruhi ekonomi di daerah tersebut, seperti kenaikan harga kayu, dan krisis keuangan [2].

Titik Panas Bumi (*hotspot*) adalah indikasi kemungkinan terjadinya kebakaran pada suatu wilayah. Titik panas (*hotspot*) merupakan daerah yang memiliki suhu permukaan relatif lebih tinggi dibandingkan daerah di sekitarnya berdasarkan ambang batas suhu tertentu yang terpantau oleh satelit penginderaan jauh. Semakin banyak titik hotspot yang terpantau oleh satelit, maka semakin banyak pula potensi terjadinya kebakaran pada suatu daerah [3]. Pada Penelitian yang dilakukan oleh (Nabila et al., 2020) mengenai penerapan algoritma K-Means Clustering untuk menentukan daerah rawan kebakaran, mengelompokan titik panas di provinsi Kalimantan Barat dengan melakukan metode data mining karena metode ini dapat mengolah data hotspot menjadi informasi yang dapat menginformasikan daerah rawan hotspot, dalam proses pengelompokan atau clustering menggunakan K-Means yaitu dengan mengelompokan berdasarkan kesamaan karakteristik data. Hasil dari K-Means clustering terdapat 3 cluster, kemudian hasil tersebut dilakukan evaluasi dengan menggunakan metode Davies-Bouldin Index (DBI) dengan nilai 3,112 yang menunjukkan bahwa hasil dari 3 cluster belum optimal [3].

Pada penelitian [4] dilakukan pemetaan kebakaran hutan yang ada di Brazil. Dataset yang digunakan adalah kebakaran hutan di Brazil dari tahun 1996 sampai 2015 yang dibagi menjadi 18 distrik dengan cakupan area sebesar 90.000 Km². Perhitungan dilakukan dengan menggunakan distribusi spasial temporal yang dibagi berdasarkan tipenya yaitu jumlah kejadian (TOc), jumlah kebakaran (TB), dan penyebab kebakaran (TIF). Pada dataset yang digunakan terdapat 516,577 total kebakaran yang terjadi (TOc), dari total data yang didapat menghasilkan 94.4% (487,503 kejadian) data yang diklasifikasikan sebagai kebakaran (TB). terdapat dua jenis kebakaran (TB) yang dipertimbangkan yaitu kurang dari 1 Ha dengan persentase sebesar 77.8% (379,261 kejadian), dan lebih dari 1 Ha, dengan persentase sebesar 22.2% (108,242 kejadian), dan data yang teridentifikasi kebakaran (TIF) sebanyak 45,603 kejadian (42.1%).

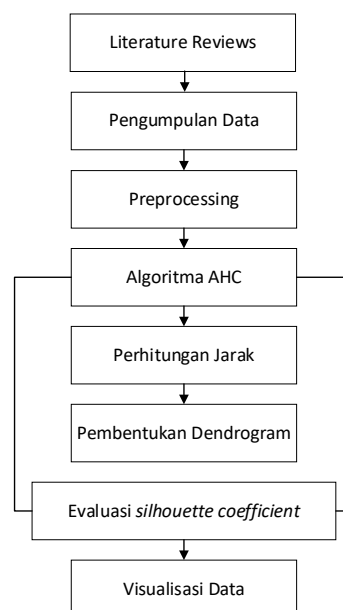
Pada penelitian yang dilakukan oleh [5] melakukan penelitian Pengelompokan Spatio-Temporal untuk Terjadinya Titik Panas Kebakaran Di Pulau Sulawesi, dalam menentukan titik panas menggunakan indikator sensor Modis dengan satelit Terra/Aqua Penelitian terkait proses identifikasi pola sebaran hotspot di suatu wilayah di Sulawesi masih terbatas dan memiliki kebaruan tersendiri. implementasi metode clustering spatio-temporal dengan algoritma ST-DBSCAN, dibantu dengan pengembangan sistem Waterfall. Hasil penelitian menunjukkan bahwa ada beberapa jenis pola distribusi yang dihasilkan oleh sistem (Stationary, Regular Reaping, Irregular, Occasional, dan Track) dari data yang digunakan dari tahun 2016 hingga 2018 di seluruh pulau Sulawesi. Dalam penelitian ini, kami juga menemukan bahwa antara

tahun 2016 hingga 2018, kemunculan titik panas dengan kecenderungan tinggi terjadi pada bulan Juli, Agustus, September, dan Oktober.

Dalam menanggulangi kebakaran hutan sendiri, antisipasi yang dapat dilakukan untuk membantu menangani pencegahan kebakaran hutan salah satunya adalah melakukan pengelompokkan terhadap wilayah yang memiliki potensi kebakaran. Penelitian ini akan melakukan Clusterisasi titik panas (*hotspot*) untuk membagi wilayah yang berpotensi terjadi kebakaran untuk segera ditindaklanjuti lebih optimal lagi. Pengelompokkan wilayah dilakukan berdasarkan cluster kelas kebakaran rendah, sedang, dan tinggi. Clusterisasi wilayah dilakukan menggunakan algoritma *Agglomerative Hierarchical Clustering (AHC)*. *Agglomerative hierarchical clustering (AHC)* dengan menggunakan bottom-up, dimulai dari masing-masing data sebagai sebuah cluster, kemudian secara rekursif mencari kelompok terdekat sebagai pasangan yang kemudian akan digabungkan menjadi kelompok yang lebih besar. Proses tersebut dilakukan berulang sampai jumlah cluster 1, sehingga akan tampak bergerak ke atas membentuk hirarki. Keunggulan dari algoritma AHC adalah tidak perlu menentukan jumlah klaster yang diinginkan [3]. Metode ini memiliki tahapan perhitungan dimulai dari pengumpulan data, pembagian kelas, perhitungan jarak menggunakan euclidean, pembentukan dendrogram, dan evaluasi yang dilakukan menggunakan *silhouette coefficient*. Data sampel titik yang digunakan pada perhitungan ini menggunakan wilayah Kalimantan Barat dengan menggunakan variable *longitude*, *latitude*, *frp*, *confidence*, dan curah hujan untuk menentukan clustering wilayah kebakaran.

2 METODE PENELITIAN

Metode penelitian menjelaskan alur penelitian secara umum yang dilakukan peneliti agar mampu menghasilkan penelitian yang sistematis guna mendapatkan hasil akhir berupa clusterisasi titik panas (*hotspot*) di wilayah provinsi Kalimantan Barat yang sesuai dengan tujuan dilakukannya penelitian ini. Adapun tahapan pada penelitian ini dapat digambarkan pada Gambar 1.



Gambar 1 Alur Methodologi Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan titik *hotspot* pada wilayah Provinsi Kalimantan Barat sejumlah 629 baris data yang diperoleh dari website NASA dengan menggunakan satellite MODIS untuk wilayah Kalimantan Barat pada tahun 2020. Terdapat 12 variabel yang ada pada dataset, namun pada penelitian ini hanya menggunakan 5 variabel yaitu data *Longitude*, *Latitude*, *Frp*, *Confidence*, Kelembapan, Curah hujan dan variable Kecepatan Angin. Variabel *confidence* digunakan sebagai variable penentu karena *confidence* memiliki nilai kepercayaan pada tingkatan panas apakah suatu wilayah mengalami kebakaran. *Confidence* adalah cakupan luas titik panas pada area 1 km² dengan satuan persen (%) [6]. Untuk melakukan perhitungan AHC, tahap awal yang harus dilakukan adalah membentuk kelas cluster terlebih dahulu [7]. Kelas cluster berperan untuk membantu pengelompokkan kategori kebakaran yang terjadi. Kolom *Cluster* pada tabel 1 masih diberi keterangan X karena Cluster akan terbentuk otomatis setelah perhitungan algoritma sudah selesai. Pembagian kelas kepercayaan dapat diperhatikan pada tabel 1 berikut

Tabel 1 Pembagian Kelas Kepercayaan

Confidence	Kelas	Cluster	Keterangan Kebakaran
0 - 45	Rendah	Cluster X	Perlu diperhatikan
46 - 85	Sedang	Cluster X	Waspada
86 - 100	Tinggi	Cluster X	Segera Tindak lanjut

2.2 Preprocessing

Tahapan preprocessing data dilakukan untuk membersihkan data sehingga data yang di digunakan tidak mengalami duplicate data. Dalam perhitungan menggunakan algoritma AHC akan menggunakan variable, *Frp*, *Confidence*, Kelembapan, Curah hujan dan variable Kecepatan angin. Variabel *Latitude* dan *longitude* akan digunakan sebagai variable untuk melakukan visualisasi data setelah cluster titik *hotspot* sudah terbentuk. Dataset yang sudah melalui proses preprocessing tidak mengalami duplikasi data sehingga data yang digunakan terdapat pada Tabel 2 berikut.

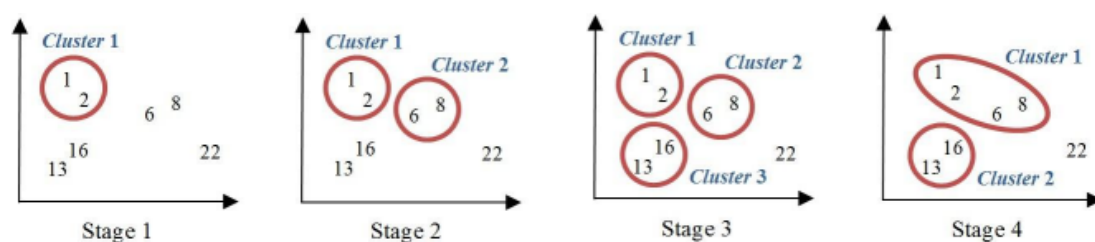
Tabel 2 Dataset

No	Longitude	Latitude	Frp	Confidence	Kelembapan	Curah Hujan	Kecepatan angin
1	109.298.393	-0.896691	6.2	55	91	9	210
2	109.687.859	-0.883583	5	54	91	9	210
3	109.490.456	-0.856683	4.7	42	91	9	210
4	109.483.643	-0.849454	6.7	49	91	9	210
5	109.793.198	-0.840304	14.5	48	91	9	210
..
361	109.382.156	-0.078379	5.7	66	91	28	190
362	10.941.478	-0.07823	8.9	71	91	28	190
363	109.381.767	-0.077027	11.8	75	91	28	190
364	10.938.166	-0.076503	18.9	82	91	28	190
365	109.583.229	-0.075438	14.8	79	91	28	190
..
626	109.315	-0.167637	4.5	57	82	0	20
627	109.301	-0.169738	7.9	59	82	0	20
628	109.342	-0.156786	109	99	82	0	20
629	109.318	-0.099562	122.1	100	82	0	20

2.3 Agglomerative Hierarchical Clustering (AHC)

Tahapan selanjutnya adalah melakukan proses *clustering*. *Hierarchical Clustering (HC)* telah banyak digunakan untuk melakukan analisis cluster karena dapat memvisualisasikan struktur hierarki cluster dengan menggunakan dendrogram. Pada umumnya *HC* terbagi menjadi dua jenis, yaitu *Agglomerative* dan *Devisve* [8] Namun, pada penelitian ini menggunakan metode *Agglomerative Hierarchical Clustering (AHC)*, dikarenakan AHC memiliki karakteristik seperti hasil clustering yang dapat dilihat dengan dendrogram, tidak perlu menentukan jumlah cluster awal, dan menggunakan pendekatan (*bottom-up*) dimana clustering data dilakukan dari kecil ke besar [9].

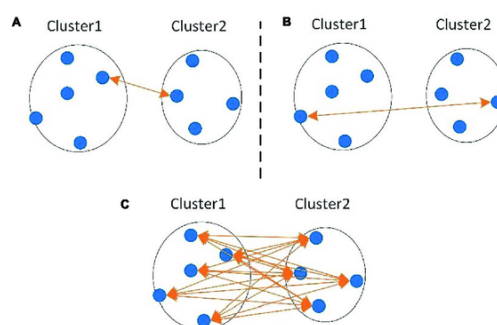
Secara garis besar AHC awalnya menetapkan cluster untuk setiap objek, kemudian mencari pasangan cluster yang paling mirip dan menggabungkannya menjadi satu cluster. Proses ini dilakukan terus menerus hingga semua objek ditempatkan pada masing-masing cluster yang paling cocok [10]. Seperti yang dapat dilihat pada Gambar 2 berikut:



Gambar 2 Ilustrasi Proses Algoritma AHC

2.4 Perhitungan Jarak

Dalam penerapannya untuk menghitung jarak clustering dari algoritma agglomerative hierarchy clustering terbagi menjadi beberapa jenis. Single-linkage adalah pengelompokan hirarki dengan sistem kerja bottom-up, sitem kerja ini dimulai dari cluster. Ilustrasi dalam penerapannya, jika pendekatan single linkage digunakan maka kita menggabungkan kedua cluster A dan B tersebut jika jarak antara setiap item di A dan item di B adalah minimum, dalam ilustrasi ini pada awalnya semua cluster dianggap terpisah dan tidak dikelompokkan, lalu pada setiap langkah iteratif dengan jarak minimum dikelompokkan bersama sehingga menjadi satu kelompok dari hasil jarak kedekatannya [11]. Complete-linkage juga dikenal sebagai tetangga terjauh. kedua cluster dapat digabungkan. Oleh karena itu, antara cluster pertama dan cluster kedua harus memiliki jarak yang jauh [12]. Average-linkage yaitu metode clustering yang mengelompokkan suatu dokumen atau objek dengan jarak rata-rata [13]. Pada gambar 3 terlihat ilustrasi dari single linkage dalam proses menentukan clustering pada ilustrasi A, Complete-linkage terlihat dalam ilustrasi B, Average-linkage terlihat dalam ilustrasi C [14].



Gambar 3 Ilustrari Perhitungan Jarak Single Linkage

Untuk menghitung jarak koefisien single linkage yaitu dengan rumus seperti berikut[15].

$$\alpha(i) + \alpha(j) = 0.5, \beta = 0, \gamma = -0.5 \quad (1)$$

Untuk menghitung jarak dengan Euclidean yaitu dengan rumus seperti berikut:

$$D_{Euc}(u, v) = \sqrt{\sum_{i=1}^n (u_i, v_i)} \quad (2)$$

Dimana:

u dan v = dua objek yang akan dihitung jaraknya

ui = komponen ke i dari u secara berurutan

2.5 Pembentukan Dendrogram

Dendrogram adalah bentuk dari pohon biner yang biasanya digunakan untuk memvisualisasikan hubungan dari hierarki dalam data seperti pada clustering. Sebuah dendrogram sering diintegrasikan dengan matriks yang dapat diurutkan ulang, yang disebut sebagai tampilan dendrogram-matriks, tetapi skala untuk menentukan lokasi jarak antar cluster sangat sulit dilakukan [16]. Algoritma ini didasarkan pada penilaian kesamaan nilai atribut mereka, pohon dendrogram sebagai diagram analisis cluster digunakan dalam pembelajaran mesin untuk kasus di mana kelas tidak ditentukan sebelumnya dan harus dikenali oleh algoritme analisis statistik dengan mencari kesamaan terdekat [17].

2.6 Evaluasi Silhouette Coefficient

Satu cara yang dapat digunakan untuk mengetahui kekuatan kluster dan melihat kualitasnya adalah menggunakan evaluasi silhouette coefficient [13]. Metode ini banyak digunakan untuk memvalidasi kluster yang menggabungkan nilai kohesi dan separasi [18]. Rentang nilai SI adalah -1 hingga +1. Nilai SI mendekati 1 menunjukkan bahwa data tersebut tidak tepat berada pada kluster tersebut. SI bernilai 0 atau mendekati 0 maka posisi data berada pada perbatasan dua kluster [18]. Rumus *Silhouette Coefficient* adalah sebagai berikut.

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i) - b(i)\}} \quad (3)$$

Dimana:

A = banyaknya data cluster A

Nilai s(i) berada antara -1 dan 1, di mana setiap nilai diinterpretasi sebagai berikut:

$s(i) \approx 1 \Rightarrow$ data ke-i digolongkan dengan baik (dalam A)

$s(i) \approx 0 \Rightarrow$ data ke-i berada di tengah antara dua kluster (A dan B)

$s(i) \approx -1 \Rightarrow$ data ke-i digolongkan dengan lemah (dekat ke kluster B daripada A)

Penafsiran nilai Silhouette coefficient ditunjukkan dalam Tabel 3 [19].

Tabel 3 Interpretasi Nilai <i>Silhouette Coefficient</i>	
<i>Silhouette Coefficient</i>	Interpretasi
0.71 - 1.00	Struktur yang dihasilkan kuat
0.51 - 0.70	Struktur yang dihasilkan baik
0.26 - 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak tersruktur

3 HASIL DAN PEMBAHASAN

Untuk memperjelas detail dalam perhitungan dapat diperhatikan pada tahapan-tahapan perhitungan dibawah ini.

3.1 Hasil Perhitungan Jarak Euclidean

Pada penelitian ini perhitungan jarak dilakukan dengan menggunakan metode Euclidean, dengan rumus sebagai berikut [20].

$$D = \sqrt{\sum_{i=1}^n (b_i - a_i)^2}$$

Selanjutnya dilakukan perhitungan dengan menggunakan seluruh data dijadikan sebagai variable pendukung. Pada perhitungan dibawah ini dilakukan dengan menggunakan 3 variabel data, yaitu *frp*, *confidence*, dan kelembapan. Sehingga diperoleh hasil sebagai berikut:

$$\begin{aligned} D_{man} &= \sqrt{(data1 - data1)^2} = \sqrt{(32 - 32)^2 + (67 - 67)^2 + (9 - 9)^2} = 0 \\ D_{man} &= \sqrt{(data1 - data3)^2} = \sqrt{(32 - 53)^2 + (67 - 94)^2 + (9 - 23)^2} = 946 \\ D_{man} &= \sqrt{(data1 - data3)^2} = \sqrt{(32 - 97)^2 + (67 - 85)^2 + (9 - 0)^2} = 470 \\ &\vdots \\ D_{man} &= \sqrt{(data30 - data28)^2} = \sqrt{(122 - 31)^2 + (100 - 87)^2 + (0 - 9)^2} = 260 \\ D_{man} &= \sqrt{(data30 - data29)^2} = \sqrt{(122 - 45)^2 + (100 - 86)^2 + (0 - 9)^2} = 273 \\ D_{man} &= \sqrt{(data30 - data30)^2} = \sqrt{(122 - 122)^2 + (100 - 100)^2 + (9 - 9)^2} = 0 \end{aligned}$$

Setelah mendapatkan hasil perhitungan jarak dari seluruh data yang dimiliki, langkah selanjutnya ialah memindahkan hasil perhitungan kedalam bentuk tabel matrix seperti yang ditunjukkan pada Tabel 4 dan 5.

Tabel 4 Matrix perhitungan jarak

Data	1	2	3	28	29	30
1	0	946	470	482	455	1260
2		0	654	600	601	634
...				0
29								0	273
30									0

Tabel 5 Matrix perhitungan jarak

Data	8	9	10	17	26	27	28
8	0	221	84	6	2	78	116
9		0	557	171	195	97	25
17				0	2	60	82
...					0
26							0	68	98

Dapat dilihat pada Tabel 4 dan 5 merupakan hasil perhitungan jarak yang telah dimasukkan kedalam tabel matrix. Dimana pada table 5 merupakan detail lanjutan dari data pada table 4. Kemudian langkah selanjutnya ialah menentukan jarak terdekat atau nilai terkecil yang

dimiliki oleh seluruh data. Tabel 6, nilai terkecil dimiliki oleh data 8 ke 26 dan 17 ke 26, dengan masing-masing bernilai 2. Oleh karena itu data 8,17, dan 26 akan digabungkan menjadi satu cluster. Tahapan tersebut terus dilakukan hingga perhitungan antara jarak data secara keseluruhan selesai dilakukan

3.2 Hasil Perhitungan Jarak Single Linkage

Langkah berikutnya ialah kembali melakukan perhitungan jarak terdekat, namun kali ini menggunakan Algoritma AHC dengan metode single linkage, dengan rumus sebagai berikut [21].

$$Min(D_{man}) = Min(d_{8,26,17}) \dots = \dots$$

Seperti yang sudah dijelaskan sebelumnya, bahwa data 8,17, dan 26 telah digabung menjadi 1 cluster, maka perhitungan dengan menggunakan metode single linkage dilakukan sebagai berikut:

$$\begin{aligned} Min(d_{8,26,17})_1 &= Min\{d_{18}, d_{126}, d_{117}\} = Min(198, 220, 244) = 198 \\ Min(d_{8,26,17})_2 &= Min\{d_{28}, d_{226}, d_{217}\} = Min(856, 824, 794) = 794 \\ Min(d_{8,26,17})_3 &= Min\{d_{38}, d_{326}, d_{317}\} = Min(146, 132, 120) = 120 \\ &\vdots \\ Min(d_{8,26,17})_{28} &= Min\{d_{288}, d_{2826}, d_{2817}\} = Min(116, 98, 82) = 82 \\ Min(d_{8,26,17})_{29} &= Min\{d_{298}, d_{2926}, d_{2917}\} = Min(111, 95, 81) = 81 \\ Min(d_{8,26,17})_{30} &= Min\{d_{308}, d_{3026}, d_{3017}\} = Min(636, 592, 550) = 550 \end{aligned}$$

Pada perhitungan diatas merupakan perhitungan dengan menggunakan metode single linkage. Perhitungan pertama dilakukan kepada data ke-1. Dimana minimal jarak/nilai dari data 1 ke 8 ialah 198, kemudian dari data 1 ke 26 ialah 220, dan data 1 ke 17 ialah 244. Selanjutnya dari ketiga data yang sudah didapatkan maka hanya data terkecil yang digunakan. Sehingga pada perhitungan data ke-1, nilai terkecil yang akan digunakan ialah 198. Begitu seterusnya perhitungan dilakukan hingga ke seluruh data.

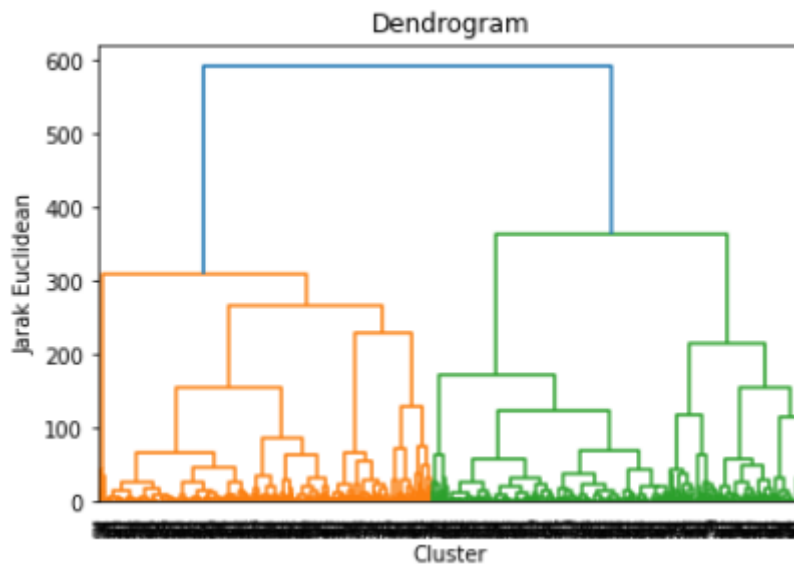
Langkah selanjutnya ialah kembali memindahkan hasil perhitungan jarak kedalam bentuk tabel matrix seperti yang dapat dilihat pada Tabel 6. Kemudian dari tabel matrix tersebut kembali akan dicari jarak terdekat untuk masing masing data, sehingga akan memunculkan *cluster* baru atau bahkan adanya data yang bergabung dengan *cluster* sebelumnya. Perhitungan akan terus dilakukan hingga seluruh data menemukan *cluster* yang cocok dengannya.

Tabel 6 Matrix perhitungan jarak

Data	{8,26,17}	9	10	18	27	28	29
{8,26,17}	0	6	84	72	60	82	81
...					0
...						0
27							0	42	47
28								0	15
29									0

3.3 Hasil Pembentukan Dendrogram

Hasil dari perhitungan jarak menggunakan perhitungan single *Euclidean* dan *Single Linkage* untuk mencari jarak kedekatan dari seluruh data yang dihitung, diperoleh pembentukan 2 *cluster* dalam bentuk dendrogram pada gambar 4 dibawah ini. Pembentukan dendrogram dapat dilakukan dengan memanfaatkan tools pada Xlstat ataupun pemrograman python untuk mempermudah pemahan pembagian *cluster* secara grafik



Gambar 4 Dendrogram Terbentuknya *Cluster*

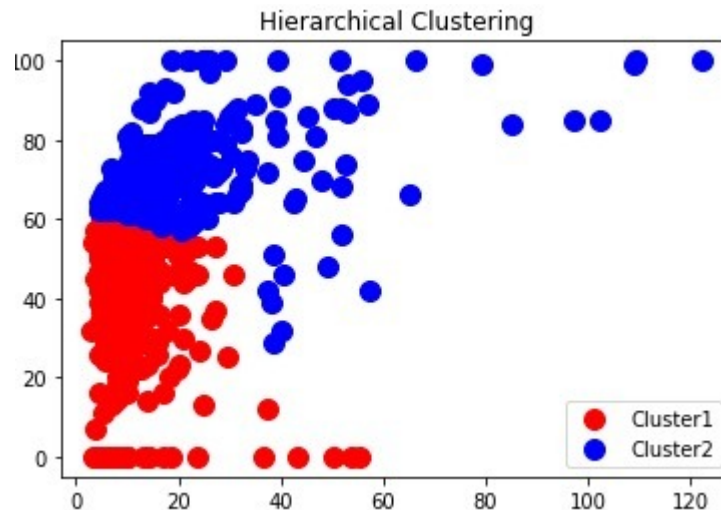
3.4 Hasil Evaluasi Shilhouette Coefficient

Berdasarkan hasil terbentuknya dendrogram, dapat diperhatikan bahwa cluster yang terbentuk berjumlah 2 cluster dengan ketentuan terbagi menjadi cluster kelas kebakaran sedang dan tinggi. Penentuan hasil kelas *cluster* masuk kedalam kebakaran sedang dan tinggi berdasarkan hasil perhitungan rata-rata dari *confidence* keanggotaan cluster 1 dan cluster 2. Nilai rata-rata *confidence* keanggotaan setiap cluster terdapat pada tabel 7. Dengan kembali merujuk pada tabel 1 untuk menentukan kelas keterangan kebakaran, maka hasil pengelompokkan cluster adalah sebagai berikut

Tabel 7 Nilai rata-rata *confidence* keanggotaan *Cluster*

<i>Cluster</i>	<i>Confidence</i> (rata-rata setelah proses hitung)	<i>Confidence</i> (acuan)	Kelas	Keterangan Kebakaran
Cluster 1	54,602	46 - 85	Sedang	Waspada
Cluster 2	82,687	86 - 100	Tinggi	Segera Tindak lanjut

Untuk mempermudah visualisai persebaran *cluster hotspot* yang terbentuk dapat diperhatikan pada gambar 5 berikut dengan ketentuan titik merah merupakan cluster 1, dan titik biru merupakan cluster 2.

Gambar 5 Sebaran Titik *Cluster* yang terbentuk

Setelah tahap persebaran titik cluster terbentuk perlu dilakukan evaluasi untuk menentukan kualitas hasil pengelompokkan masing-masing perhitungan jarak yang sudah dilakukan dengan beberapa kondisi scenario perhitungan kombinasi variable yang digunakan pada tabel 6

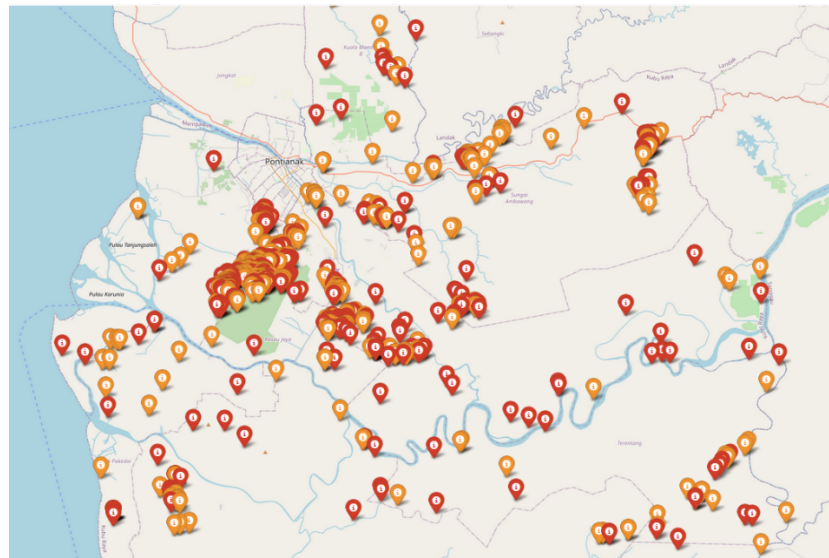
Tabel 9 Evaluasi *Silhouette Coefficient*

Skenario	Variabel	Jumlah Cluster			
		2	3	4	5
1	<i>frp, confidence</i> , kelembapan, curah				
	hujan, kecepatan angin	0,771	0,753	0,752	0,667
2	<i>frp, confidence</i> , kelembapan, curah hujan	0,626	0,555	0,538	0,452
3	<i>frp, confidence</i> , kelembapan	0,808	0,675	0,669	0,556

3.5 Hasil Visualisai Pemetaan

Hasil dari nilai evaluasi *Silhouette Coefficient* menunjukkan bahwa dengan dilakukan beberapa scenario pengujian dengan beberapa variable disetiap uji coba, nilai *Silhouette Coefficient* pada cluster 2 memiliki nilai yang dominan tinggi sehingga value kualitas pada terbentuknya 2 cluster adalah baik. Sehingga selanjutnya dapat dilakukan visualisasi pada peta menggunakan variable longitude dan latitude berdasarkan persebaran cluster yang terbentuk. Visualisasi pada peta menggunakan maps leaflet.

Leaflet adalah salah satu perpustakaan JavaScript open-source paling populer untuk peta interaktif [22]. Proses dalam visualisasi menggunakan library dari folium untuk bisa menggunakan leaflet map. Pertama, buat function location untuk mentukan lokasi dari setiap data menggunakan dataframe dari (Lat, Long). Lalu membuat looping menggunakan aturan if else, dengan menggunakan dataframe dari dendrogram yang terbentuk sebagai acuan untuk menentukan nilai if, jika nilai dendrogram kurang dari satu maka akan masuk kedalam kelompok dalam pengawasan dengan indikator warna orange dan selain itu membutuhkan bantuan segera dengan indikator merah. Informasi yang terdapat disetiap titik memiliki dataframe *longitude*, *latitude*, *frp*, *confidence*, curah hujan, kelembapan, dan kecepatan angin. Untuk hasil visualisasi dapat diperharikan pada gambar 6 berikut



Gambar 6 Visualisasi persebaran titik *hotspot* Maps Leaflet

4 KESIMPULAN

Dari hasil pembahasan dan analisis yang sudah dilakukan, dapat ditarik kesimpulan sebagai berikut:

1. Pembentukan kelas cluster terbagi kedalam 3 kelas, kecil dengan range panas (0-0), sedang dengan range (0-0), dan bahaya dengan range (0-0) yang dapat disesuaikan dengan keperluan penelitian.
2. Algoritma *Agglomerative Hierarchical Clustering (AHC)* pada penelitian ini mampu bekerja dengan pengujian evaluasi terbaik dengan *Silhouette Coefficient* lebih baik pada terbentuknya 2 cluster, yaitu pada skenario 1 sebesar 0,771 dan skenario 3 sebesar 0,808.
3. Hasil dari perhitungan ini masih dapat ditingkatkan dengan memperhatikan faktor eksternal pendukung seperti nilai GDP yang dapat menjadi pertimbangan pola perilaku masyarakat akan perlindungan lahan tanam.
4. Clusterisasi titik panas yang terbentuk dapat menjadi lokasi yang memerlukan perhatian khusus untuk pemantauan lebih lanjut kelokasi titik panas. Selaras dengan terbentuknya titik api, pembentukan pos pantau dapat dipertimbangkan lokasinya dengan memanfaatkan hasil cluster titik panas yang dihasilkan.

5 SARAN

Pada penelitian berikutnya diharapkan untuk menggunakan dataset yang lebih banyak lagi dengan memperhatikan variable eksternal dan internal seperti variabel *bright* sebagai indikator baru dalam perhitungan untuk memprediksi besaran kebakaran lahan hutan.

6 UCAPAN TERIMA KASIH

Penulis mengucapkan terimakasih yang sebesar-besarnya disampaikan penulis kepada Universitas Amikom Yogyakarta sebagai pihak telah memberi dukungan financial terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] BNPB, “Kerugian Kebakaran Hutan dan Lahan Sepanjang 2019 Capai Rp 75 Triliun - BNPB,” 2019. <https://bnpb.go.id/berita/kerugian-kebakaran-hutan-dan-lahan-sepanjang-2019-capai-rp-75-triliun> (accessed Jul. 25, 2022).
- [2] S. Sumaryati, N. Cholianawati, and A. Indrawati, “The impact of forest fire on air-quality and visibility in Palangka Raya,” *Journal of Physics: Theories and Applications*, vol. 3, no. 1, p. 16, Mar. 2019, doi: 10.20961/jphystheor-appl.v3i1.38071.
- [3] 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.
- [4] A. C. Meira Castro, A. Nunes, A. Sousa, and L. Lourenço, “Mapping the causes of forest fires in Portugal by clustering analysis,” *Geosciences (Switzerland)*, vol. 10, no. 2, Feb. 2020, doi: 10.3390/geosciences10020053.
- [5] A. Vatesia, R. R. Rais, F. P. Utama, and W. Oktarianti, “MINING FIRE HOTSPOTS OVER NUSA TENGGARA AND BALI ISLANDS,” *Indonesian Journal of Forestry Research*, vol. 9, no. 1, pp. 73–85, 2022, doi: 10.20886/ijfr.2022.9.1.73-85.
- [6] N. Huda, N. Debataraja, and E. Sulistianingsih, “ESTIMASI CONFIDENCE TITIK PANAS PADA KEBAKARAN HUTAN MENGGUNAKAN METODE ORDINARY KRIGING,” doi: 10.13140/RG.2.2.35901.87526.
- [7] O. Yim and K. T. Ramdeen, “Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data,” 2015.
- [8] S. Hirano, X. Sun, and S. Tsumoto, “Comparison of clustering methods for clinical databases,” *Information Sciences*, vol. 159, no. 3–4, pp. 155–165, Feb. 2004, doi: 10.1016/j.ins.2003.03.011.
- [9] F. Murtagh and P. Contreras, “Algorithms for hierarchical clustering: An overview,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 2, no. 1, pp. 86–97, Jan. 2012, doi: 10.1002/widm.53.
- [10] E. N. Narciso, M. Delamaro, and F. L. S. Nunes, “Test case selection using CBIR and clustering Test Case Selection Using CBIR View project Automation of test Oracles for the processing of medical images of three-dimensional models View project,” 2013. [Online]. Available: <https://www.researchgate.net/publication/290332664>
- [11] Institute of Electrical and Electronics Engineers and Manav Rachna International Institute of Research and Studies, *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing : trends, prespectives and prospects : COMITCON-2019 : 14th-16th February, 2019*.
- [12] P. Phetsangkat, K. Chalermyanont, and R. Duangsoithong, “Hierarchical Clustering Electric Load: Case Study in Lower South Region of Thailand,” in *2019 16th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, Jul. 2019, pp. 881–884. doi: 10.1109/ECTI-CON47248.2019.8955425.

- [13] Y. Reinaldi, N. Ulinuha, and Moh. Hafiyusholeh, "Comparison of Single Linkage, Complete Linkage, and Average Linkage Methods on Community Welfare Analysis in Cities and Regencies in East Java," *Jurnal Matematika, Statistika dan Komputasi*, vol. 18, no. 1, pp. 130–140, Sep. 2021, doi: 10.20956/j.v18i1.14228.
- [14] Z. G. Wei, X. D. Zhang, M. Cao, F. Liu, Y. Qian, and S. W. Zhang, "Comparison of Methods for Picking the Operational Taxonomic Units From Amplicon Sequences," *Frontiers in Microbiology*, vol. 12, Mar. 2021, doi: 10.3389/fmicb.2021.644012.
- [15] NCSS and LLC, "445-1 Hierarchical Clustering / Dendrograms."
- [16] M. Forina, C. Armanino, and V. Raggio, "Clustering with dendrograms on interpretation variables," *Analytica Chimica Acta*, vol. 454, no. 1, pp. 13–19, Mar. 2002, doi: 10.1016/S0003-2670(01)01517-3.
- [17] P. Lemenkova, "R Libraries {dendextend} and {magrittr} and Clustering Package scipy.cluster of Python For Modelling Diagrams of Dendrogram Trees," *Carpathian Journal of Electronic and Computer Engineering*, vol. 13, no. 1, pp. 5–12, Sep. 2020, doi: 10.2478/cjece-2020-0002.
- [18] G. P. Trayasiwi, "PENERAPAN METODE KLASSTERING DENGAN ALGORITMA K-MEANS UNTUK PREDIKSI KELULUSAN MAHASISWA PADA PROGRAM STUDI TEKNIK INFORMATIKA STRATA SATU."
- [19] A. Struyf, M. Hubert, and P. J. Rousseeuw, "Clustering in an Object-Oriented Environment."
- [20] K. K. Mohbey and G. S. Thakur, "An Experimental Survey on Single Linkage Clustering," *International Journal of Computer Applications*, vol. 76, no. 17, pp. 6–11, Aug. 2013, doi: 10.5120/13337-0327.
- [21] D. T. Utari and D. S. Hanun, "Hierarchical Clustering Approach for Region Analysis of Contraceptive Users," *EKSAKTA: Journal of Sciences and Data Analysis*, pp. 99–108, Sep. 2021, doi: 10.20885/eksakta.vol2.iss2.art3.
- [22] P. K. Litoria, T. Singh Bhatia, H. Singh, P. K. Litoria, and B. Pateriya, "Web GIS Development using Open Source Leaflet and Geoserver Toolkit," vol. 9, [Online]. Available: <https://www.researchgate.net/publication/326971834>