Jurnal Publikasi Ilmu Komputer dan Multimedia

E-ISSN: 2808-8999
P-ISSN: 2808-9375

(Artikel Penelitian/Ulasan)

Analisa Daerah Kebakaran Hutan Pulau Sumatera dengan K-Means Clustering Metode CRISP-DM

Winarnie [1]*, Hery Oktafiandi[2], Pebriyanti Panjaitan [3], M. Fajar Ramadhan [4], dan Yohanes [5]

[1,3,5] Teknik Informatika, Satu University, Jl. BKR No.63, Ancol, Kec. Regol, Kota Bandung, Jawa Barat
[2,4] Sistem Informasi, Satu University, Jl. BKR No.63, Ancol, Kec. Regol, Kota Bandung, Jawa Barat
* Penulis : winarnie@univ.satu.ac.id

Abstract: Forest and Land Fires (Karhutla) often occur in Indonesia. The 2015 forest and land fires in Sumatra and Kalimantan have become a global concern because they have brought haze pollution to neighboring countries. Information about the grouping of areas affected by forest fires is used to remind people to be more alert and prepared. The K-Mean Clustering method with the CRISP-DM data mining model can be used as a way to analyze areas that are prone to forest and land fires. This study was conducted to create groups of areas with a high probability of forest fires and areas that are safe from forest fire disasters in regencies on the island of Sumatra which includes 9 provinces on the island of Sumatra. The forest and land fire alert area group is represented by several clusters. In this study, it succeeded in producing good and optimal clusters as seen by the positive silhouette value and close to 1. The best value cluster is cluster 2 with the largest silhouette coefficient index value of 0.96 with a forest and land fire area of 6624 ha and the emergence of 6006 hotspots

Keywords: *Hot Spot; CRISP-DM; K-Mean; Clustering, Silhoutte Coefficient.*

Abstrak: Kebakaran Hutan dan Lahan (Karhutla) sering terjadi di negara Indonesia. Karhutla pada wilayah Sumatera dan Kalimantan tahun 2015 menjadi sorotan dunia karena telah membawa polusi kabut asap ke negara tetangga. Informasi tentang pengelompokkan daerah yang terkena dampak kebakaran hutan digunakan untuk mengingatkan supaya masyarakat lebih waspada dan siap. Metode K-Mean Clustering dengan model data mining CRISP-DM, bisa dipakai sebagai salah satu cara untuk menganalisa daerah yang termasuk dalam daerah rawan terjadi karhutla. Penelitian ini dilakukan untuk membuat kelompok daerah dengan kemungkinan kebakaran hutan yang tinggi. dan daerah yang aman dari bencana kebakaran hutan pada kabupaten-kabupaten yang terdapat pada pulau Sumatera yang meliputi 9 provinsi di Pulau Sumatera., Kelompok daerah waspada kebakaran hutan dan lahan diwakili oleh beberapa cluster. Pada penelitian ini berhasil menghasilkan cluster yang baik dan optimal dapat dilihat dengan nilai *silhouette* bernilai positif dan mendekati 1. Cluster nilai terbaik adalah cluster 2 dengan nilai *silhouette coefficient* indeks terbesar yaitu 0,96 dengan luas karhutla sebanyak 6624 ha dan kemunculan titik api sebanyak 6006 titik.

Kata kunci: Titik Api; CRISP-DM; K-Mean; Clustering; Silhoutte Coefficient.

Diterima: 25 Oktober 2025
Direvisi: 28 Oktober 2025
Diterima: 30 Oktober 2025
Diterbitkan: 30 Januari 2026
Versi sekarang: Januari 2026

Hak cipta: © 2025 oleh penulis. Diserahkan untuk kemungkinan publikasi akses terbuka berdasarkan syarat dan ketentuan lisensi Creative Commons Attribution (CC BY SA) (https://creativecommons.org/licenses/by-sa/4.0/)

1. Pendahuluan

Indonesia merupakan negara dengan hutan tropis yang luas serta mengandungi kekayaan hayati tertinggi di dunia selain Brasil dan Kolombia. Pada tahun 2007, 2012 dan 2015 terjadi kebakaran hutan wilayah Sumatera dan Kalimantan membawa polusi udara asap lintas batas pada daerah ASEAN.[1]

Karhutla merupakan ancaman serius bagi negara kita. Kasus karhutla di Indonesia semakin meningkat setiap tahunnya. Ada 20.850 titik api pada tahun 2012. Sebagian besar kebakaran sekitar 92% pada daerah Kalimantan, Sumatera [2]. Karhutla di Indonesia disebabkan oleh beberapa factor. Faktor alam yang sering menyebabkan karhutla ialah musim kemarau yang panjang. Terjadinya karhutla memiliki faktor dan dampak yang ditimbulkan. Adapun faktor yang ditimbulkan dari bencana kebakaran hutan terbagi dalam dua kelompok yaitu, kelompok alam dan kelompok campur tangan manusia. [3] Pemicu dari terjadinya karhutla karena lapisan-lapisan organik yang sudah kering sehingga akan mudah terbakar, hal tersebut yang dapat memicu sering terjadinya kebakaran hutan dan lahan. [4]. Mengetahui adanya faktor terjadinya karhutla, adalah penting untuk dijadikan pencegahan karhula sejak dini. [5]

Metode clustering adalah metode dalam data maining yang bertujuan mengklasterkan data, data yang karakteristi mirip diklasterkan pada satu grup yang sama [6], analisa cluster dengan menggunakan k-means clustering memiliki kelebihan tidak sesnsitif pada outlier dan dapat mengurangi noise [7]. K-means clustering merupakan teknik clustering yang populer [8] karena metode k-means clustering mudah digunakan [9]. Kelebihan lain dari k-means clustering yaitu dapat mengklasterkan data dengan jumlah dan variabel yang besar serta dipakai pada skala data ordinal, interval dan rasio [10].

Penelitian ini memakai teknik CRISP-DM untuk analisa masalah yang ada pada percobaan ini. Metode ini ada 6 proses yaitu : business understanding, data understanding, data preparation, modelling, evaluation, dan deployment [11].

Percobaan dilakukan dengan clustering teknik k-mean clustering dengan metode CRISP-DM untuk menentukan kelompok rawan titik api di Pulau Sumatera. Hasil percobaan clustering ini dapat memberikan informasi dan pembelajaran pada masyarakat dapat mencegah terjadinya karhutla di Pulau Sumatera meliputi daerah rawan titik api dan daerah tidak rawan titik api. Diharapkan dapat meningkatkan kesadaran masyarakat turut menjaga dan mencegah terjadinya karhutla.

Beberapa pecobaan telah dilakukan berhubungan dengan karhutla maupun non karhutla menggunakan metode clustering. Percobaan tersebut menggunakan berbagai sumber dataset untuk data penelitian. Beberapa penelitian tersebut dijadikan acuan dalam penelitian ini, dan disajikan sebagai tinjauan literature

Dyang Falia P d.k.k [12] melakukan penelitian menggunakan data hotspot dengan parameter lintang, bujur, magnitude, frp(radiant flux of fire), dan confident menggunakan K-Medoids. Hasil dari penelitiannya menggunakan K-Medoids berhasil pada proses clustering data hotspot.

Penelitian lain dilakukan oleh Ahmad Harmain d.k.k [13] yang melakukan penelitian mengenai titik panas dengan metode clustering bersumber data Satelit MODIS, kesimpulannya adalah praproses bentuk normalisasi.

Easbi Ikhsan melakukan percobaan berkenaan dengan analisa perilaku anggota pelatihan kursus. Menghasilkan penelitian menjadi 3 kelompok atau cluster merunut kegiatan anggota pelatihan. [14]

Penelitian yang berhubungan dengan clustering kebakaran hutan juga dilakukan oleh Arthifaturrotitah d.k.k [15] dan Anggi Ayu D.S d.k.k [16] . Penelitian [15] membandingkan pengelompokan KMeans dan K-Medoids untuk pengklasteran kemungkinan kebakaran hutan/lahan menurut sebaran hotspot di Indonesia pada tahun 2018. Sehingga diperoleh hasil Silhouette Coefficient pada metode KMeans adalah 0,558. Nilai Silhouette Coefficient menggunakan cara K-Medoids adalah 0,529 disimpulkan dengan metode K-Means mendapatkan nilai SC lebih baik dibandinkan K-Medoids, K-Means mampu memberikan hasil klaster yang lebih baik. Pada penelitian [16] penggunaan teknik clustering metode algoritma K-Medoids dengan data transaksi penjualan Pengujian kelompok dengan nilai Silhouette dan nilai Davies Boulbin dengan tujuan mengetahui jumlah kelompok yang paling baik. Hasil penelitiannya adalah jumlah kelompok optimal adalah 3 (tiga) kelompok.

Berdasarkan tinjauan terhadap penelitian-penelitian sebelumnya, sebagian besar kajian terkait kebakaran hutan dan lahan masih berfokus pada pengelompokan hotspot menggunakan satu jenis data atau satu periode waktu tertentu, serta belum banyak mengaitkan secara langsung antara luas kebakaran hutan dan jumlah hotspot pada level kabupaten/kota secara terintegrasi. Selain itu, penggunaan model CRISP-DM pada studi kebakaran hutan di Pulau Sumatera masih terbatas dan umumnya hanya menekankan pada aspek pemodelan tanpa analisis interpretatif terhadap karakteristik setiap klaster yang terbentuk.

Oleh karena itu, kontribusi utama penelitian ini adalah menerapkan metode K-Means Clustering dengan kerangka CRISP-DM untuk mengelompokkan daerah rawan kebakaran hutan dan lahan di Pulau Sumatera berdasarkan kombinasi dua indikator utama, yaitu luas karhutla dan jumlah hotspot, pada level kabupaten/kota.

Kebaruan penelitian ini terletak pada:

1. pengintegrasian data luas kebakaran dan sebaran hotspot dalam satu proses clustering,
2. analisis hasil klaster yang tidak hanya berhenti pada evaluasi metrik silhouette, tetapi juga interpretasi karakteristik risiko tiap klaster, dan
3. penyajian hasil clustering sebagai dasar klasifikasi tingkat kerawanan karhutla yang dapat mendukung upaya mitigasi dan peringatan dini kebakaran hutan di wilayah Sumatera.

2. Tinjauan Literatur

2.1. Kebakaran Hutan dan Lahan (Karhutla)

Kebakaran hutan dan lahan merupakan bencana ekologis yang sering terjadi di Indonesia dan berdampak luas pada lingkungan, kesehatan, ekonomi, serta hubungan internasional. Agung et al. (2018) menjelaskan bahwa Indonesia adalah salah satu negara dengan hutan tropis terluas di dunia, sehingga tingkat kerentanan terhadap kebakaran sangat tinggi. Banyak kasus karhutla besar terjadi pada tahun 2007, 2012, dan 2015 yang menyebabkan kabut asap lintas batas hingga ke negara ASEAN.

Faktor penyebab karhutla dapat dibagi menjadi dua kelompok besar, yaitu faktor alam dan faktor manusia. Supriyanto dkk. (2018) menyatakan bahwa sebagian besar kebakaran dipicu oleh aktivitas manusia seperti pembukaan lahan. Namun, kondisi alam seperti musim kemarau panjang juga meningkatkan risiko kebakaran (Muzaki et al., 2021). Selain itu, keberadaan vegetasi organik kering turut memperbesar potensi terbakarnya suatu wilayah (Yusuf et al., 2019). Oleh karena itu, identifikasi dini daerah rawan karhutla menjadi kebutuhan penting dalam mitigasi.

2.2. Data Mining dan Metode Clustering

Data mining adalah proses penggalian pola atau pengetahuan dari data dalam jumlah besar. Salah satu teknik yang banyak digunakan adalah clustering, yaitu pengelompokan data berdasarkan kemiripan karakteristik. Handayani, Teguh, dan Lestari (2021) menjelaskan bahwa clustering berfungsi mengelompokkan data dengan karakteristik serupa ke dalam grup yang sama.

K-Means merupakan salah satu algoritma clustering yang paling populer dan efektif karena sifatnya yang sederhana, cepat, dan mampu menangani data skala besar (Vulandari et al., 2020; Gopikaramanan et al., 2015). Wibowo et al. (2021) juga menambahkan bahwa K-Means relatif tidak sensitif terhadap outlier dan dapat mengurangi noise sehingga banyak digunakan dalam penelitian berbasis data numerik.

Selain K-Means, terdapat algoritma K-Medoids yang lebih tahan terhadap outlier, namun sering kali menghasilkan kualitas cluster yang lebih rendah dibandingkan K-Means pada dataset tertentu. Studi Athifaturrofifah et al. (2019) menunjukkan bahwa nilai Silhouette Coefficient pada K-Means (0,558) lebih tinggi daripada K-Medoids (0,529), yang berarti K-Means memberikan cluster yang lebih baik.

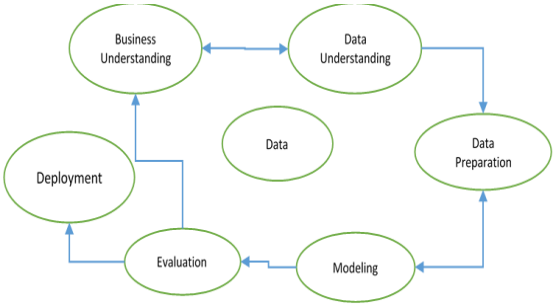
2.3. Model CRISP-DM dalam Analisis Data

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah kerangka kerja analisis data yang terdiri atas enam tahapan:

a. Business understanding,
b. Data understanding,
c. Preparation,
d. Modeling,
e. Evaluation,
f. Deployment

3. Metode

Percobaan ini menggunakan teknik CRISP-DM untuk menyelesaikan permasalahan secara garis besar pada bisnis dan penelitian yang memiliki 6 tahapan [17] Langkah data mining model crisp-dm dengan 6 tahap dapat dijabarkan sebagai berikut yaitu gambar 1.



Gambar 1. Tahapan CRISP-DM

3.1.Business Understanding

Business understanding, adalah langkah awal pada proses CRISP-DM atau dapat dinamai pengertian permasalahan penelitian. Pemahaman masalah pada penlitian ini adalah mengelompokkan atau mengkasterkan daerah rawan karhula di pulau Sumatera dengan mengacu data sebaran titik panas dan luas daerah hutan /lahan yang terbakar di wilayah pulau Sumatera. Data – data dipelajari untuk mendapatkan dataset yang sesuai dengan penelitian. Data sebaran titik api dan luas daerah kebakaran hutan di wilayah pulau Sumatera digunakan sebagai acuan dalam pengelompokkan daerah rawan bencana kebakaran hutan sehingga diharapkan kelompok dalam wilayah bahaya akan lebih waspada.

3.2. Data Understanding

Pada langkah ini adalah diawali pengumpulan data, menggambarkan data, dan menguji kualitas data. Data yang dipakai diolah dengan proses penjabaran pada setiap fiturnya. Menggunakan data sebagai berikut:

3.2.1. Data luas karhutla pada Wilayah Pulau Sumatera

Terdiri dari dataset nama kabupaten/kota dengan 6 atribut yang berisi jumlah luas hutan yang terbakar pada tahun 2017, 2018, 2019, 2020, 2021, dan 2022.

3.2.2 Dataset Sebaran Titik Api pada Wilayah Pulau Sumatera

Terdiri dari data nama provinsi dan 5 atribut yang berisi nama provinsi, nama kabupaten/kota, satelit, confidence, dan counter/jumlah titik api

3.3. Data Preparation

Langkah ini mempersiapkan dan pengecekan data sehingga data yang digunakan sesuai dengan dataset pada penelitian. Data diperbaiki sebelum masuk ke pemodelan. Pada tahap ini dilakukan beberapa tahap:

3.3.1 Data Selection

Langkah selanjutnya dilakukan adalah memilih data yang benar – benar dibutuhkan pada penelitian

3.3.2 Data prepossessing

Pada data preprosessing adalah tahap mempersiapkan data untuk siap dipakai pada tahap pemodelan. Pada tahap ini terdapat proses cleansing untuk menganalisa kualitas suatu data dengan menghilangkan data yang tidak dibutuhkan. Selanjutnya juga terdapat proses missing value yaitu mencari data yang hilang atau kosong dan menambahkan dengan angka null. Proses selanjutnya yaitu merge yaitu menggabungkan data untuk mendapatkan kolom dari tabel data yang dibutuhkan.

3.3.3 Data Transformation

Langkah ini adalah mengubah data menjadi bentuk lain sehingga bisa diproses pada langkah selanjutnya.

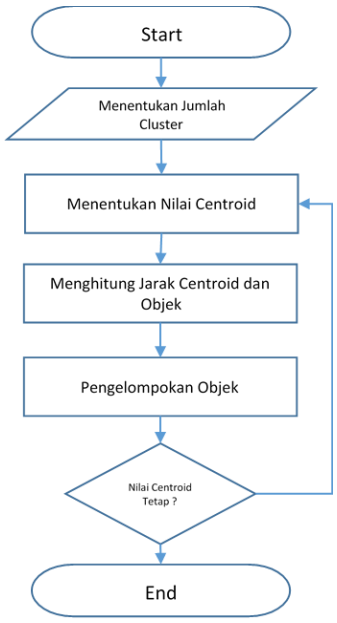
3.4. MODELLING

Tahap modelling adalah tahap menggunakan model clustering k-means, adapun langkah – langkah pada perhitungan secara rumus tahap modelling adalah :[18]

1. Menentuan jumlah kelompok

2.. Membuat Nilai Centroid
3.. Membuat perhitungan Jarak antara Centroid dan Objek
4.. Mengelompokkan Objek
5.. Balik ke langkah 2. Melakukan perhitungan sampai nilai Centroid tetap dan Objek tetap.

Gambar 2 merupakan alur dari pemodelan clustering yang digunakan pada penelitian ini selain dengan perhitungan manual rumus, penelitian ini juga menggunakan pemodelan dengan Bahasa python yaitu dengan menggunakan metode elbow.



Gambar 2. Alur Model K-Means

3.5. Evaluation

Proses selanjutnya setelah menentukan model adalah tahap evaluasi agar mendapatkan model yang paling tepat dengan menggunakan metode silhouette coefision. Silhoutte coefision berdasarkan metode penafsiran dan validasi konsistansi cluster data. Seberapa baik nilai silhouette objek dalam cluster adalah pengukuran seberapa mirip objek dalam cluster dibandingkan pada cluster lain. Langkah– langkah dalam menghitung silhouette coefision adalah:

Mencari nilai rata-rata jarak data i dengan seluruh data dalam cluster yang sama dengan Pers. (1).

$$\boldsymbol{ai} = \frac{\mathbf{1}}{\mathbf{[A]-1}} \sum \boldsymbol{j} \in \boldsymbol{A}, \boldsymbol{j} \neq \boldsymbol{d}(\boldsymbol{i},\boldsymbol{j}) \quad (1)$$

Mencari rata-rata data *i* dengan seluruh data cluster yang lain dengan Pers. (2)

$$\boldsymbol{b}(\boldsymbol{i}) = \mathbf{min}(\boldsymbol{D}(\boldsymbol{i},\boldsymbol{c})) \quad (2)$$

Mencari nilai silhouette coefisien menggunakan Pers.(3)

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

Kisaran nilai silhouette coefision antara -1 sampai 1. Nilai tersebut baik karena mempunyai pencapaian silhouette positif. Jika nilai silhouette sama dengan 1 posisi objek berada pada kelompok yang tepat, kalau pencapaiani silhouttenya adalah 0 posisi objek di stuktur kelompok yang kurang jelas, dan bila silhouttenya -1 maka stuktur clusternya mempunyai pencapaian overlapping maka posisi objek lebih baik dipindahkan kedalam cluster lain..

3.6. Deployment

Tahap ini penelitian dilakukan dengan dibuat dan diterapkan dengan membuat laporan berisi hasil yang telah dibangun oleh model sebelumnya

4. Hasil dan Pembahasan

4.1 . Business Understanding

Langkah percobaan pertama adalah berkaitan adanya bahaya kebakaran hutan yang terjadi di wilayah pulau Sumatera. Data sebaran titik panas dikawasan pulau Sumatera dan luas hutan/lahan yang terbakar di wilayah pulau Sumatera menjadi acuan padapercobaan ini. Tujuan dari percobaan adalah didapat suatu pengetahuan tambahan mengenai daerah rawan bencana karkutla sehingga dapat menjadi lebih waspada dan perduli akan terjadinya kebakaran. Batasan masalahnya yaitu daerah yang dijadikan objek penelitian ini adalah pulau Sumatera dengan mengambil data berupa sebaran titik panas dan luas hutan dan lahan yang terbakar pada wilayah pulau Sumatera.

4.2. Data Understanding

menggunakan data bersumber pada Sipongi milik kementrian lingkungan hidup dan kehutanan https://sipongi.menlhk.go.id/. Data yang diambil adalah luas wilayah karhutha pada pulau sumatra pada tabel 1 dan sebaran titik api (hotspot) pulau sumatra periode 1 Januari 2022 sampai 25 Juni 2022 ditunjukkan pada tabel 2. Dataset ini terdiri dari 149 record yang meliputi nama kab dan 6 atribut yaitu: provinsi, kab, luas karhutla, satelit, confident dan hotspot. Berikut sebagian dataset yang digunakan

Tabel 1. Data Luas Kebakaran Luas Dan Lahan

Kab/Kota	2017	2018	2019	2020	2021	2022
Tebo	12	134,00	6.390,00	102,00	104,00	12,00
Kepulauan Anambas	20	0	0	0	0	20,00
Kota Batam	0	0	1.438,00	144,00	119,00	0
Lingga	0	310,00	179,00	202,00	32,00	0
Natuna	0	11,00	3.704,00	8.001,00	380,00	0

Tabel 2. Data Sebaran Titik Api

Provinsi	Kab/Kota	Satelit	Confidence	Hotspot
Aceh	Aceh Barat	NASA-NOAA20	Low	41

Aceh	Aceh Barat Daya	NASA-NOAA20	Medium	28
Aceh	Aceh Besar	NASA-NOAA20	Low	25

4.3. **Data Preparation**

Tahapan ini yaitu menyiapkan data disesuaikan supaya dataset bisa digunakan untuk proses pemodelan.

4.3.1 Data Selection

Dataset menggunakan dataset periode 1 Januari 2022 sampai 25 Juni 2022 sebanyak 149 data. Dari total ada 10 atribut menjadi 6 atribut yang terpilih yaitu provinsi, kab, luas karhutla dan hotspot. Data hasil seleksi terlihat di tabel 3.

Tabel 3 *Dataset Preparation*

Provinsi	Kab/Kota	Luas Karhutla	Satelit	Confidence	Hotspot
Aceh	Aceh Barat	30	NASA-NOAA20	Low	41
Aceh	Aceh Barat Daya	15	NASA-NOAA20	Medium	28
Aceh	Aceh Besar	0	NASA-NOAA20	Low	25
Aceh	Aceh Jaya	99	NASA-NOAA20	Medium	68

4.3.2. Data Preprocessing

Pada preprocessing dilakukan pembenahan data, diantaranya membenahi data yang hilang. Penelitian ini memakai data sebanyak 149 data. Dari total ada 6 atribut menjadi 4 atribut yang terpilih yaitu provinsi, kab, luas karhutla dan hotspot. Data hasil seleksi dapat dilihat pada gambar 3

Kemudian selanjutnya adalah proses data transformation

	Provinsi	Kab/Kota	Luas Karhutla	Hotspot
0	Aceh	ACEH BARAT	30	41
1	Aceh	ACEH BARAT DAYA	15	28
2	Aceh	ACEH BESAR	0	25
3	Aceh	ACEH JAYA	99	68
4	Aceh	ACEH SELATAN	322	307
...	...	...	...	...
144	Sumatera Utara	TAPANULI SELATAN	246	288
145	Sumatera Utara	TAPANULI TENGAH	238	179
146	Sumatera Utara	TAPANULI UTARA	7	145
147	Sumatera Utara	TEBING TINGGI	0	2
148	Sumatera Utara	TOBA SAMOSIR	6	60

149 rows × 4 columns

Gambar 3. Data Preprocessing

4.3.3. Data Transformation

Langkah ini adalah menstandarkan ukuran data variabel dari bentuk angka pada data ke bentuk nilai 0-1 pada gambar 4. Penelitian menggunakan. Standard Scaler.

```
# Menstandarkan ukuran variabel
scaler = MinMaxScaler() #fungsinya untuk mengskalakan
x_scaled = scaler.fit_transform(X)
x_scaled
```

```
array([[4.52898551e-03, 6.82650683e-03],
       [2.26449275e-03, 4.66200466e-03],
       [0.00000000e+00, 4.16250416e-03],
       [1.49456522e-02, 1.13220113e-02],
       [4.86111111e-02, 5.11155511e-02],
       [6.00845411e-02, 3.61305361e-02],
       [0.00000000e+00, 3.33000333e-04],
       [0.00000000e+00, 1.99800200e-03],
       [0.00000000e+00, 1.21545122e-02],
       [0.00000000e+00, 7.82550783e-03],
       [0.00000000e+00, 2.83050283e-03],
       [0.00000000e+00, 1.33200133e-03],
       [0.00000000e+00, 3.82950383e-03],
       [0.00000000e+00, 6.82650683e-03],
       [0.00000000e+00, 0.00000000e+00],
       [0.00000000e+00, 0.00000000e+00],
       [0.00000000e+00, 3.39660340e-02],
       [0.00000000e+00, 3.33000333e-04],
```

Gambar 4. Data Transformation

4.4. Modelling

Langkah percobaan selanjutnya adalah teknik *clustering* dengan menggunakan bahasa pemrograman *python* dan perhitungan rumus untuk mendapatkan pola data mining. Teknik pemodelannya dengan model k-means. langkah perhitungan pemodelan k-means adalah:

4.4.1. Menentuan jumlah *cluster*

Jumlah kluster pada percobaan ini adalah 7 klaster dengan simbol c1,c2,c3,c4,c5,c6, dan c7.

4.4.2. Menentukan Nilai *Centroid*

Titik centroid yang digunakan adalah

c1 = (99,68)

c2 = (6624,6006)

c3 = (4148,204)

c4 = (988,760)

c5 = (317,563)

4.4.3. Menghitung Jarak antara *Centroid* dan Objek

Melakukan perhitungan jarak antar centroid dan objek:

d0 (30,41) dengan c1(99,68)

$$d0 = \sqrt{(30-99)^2 + (41-68)^2} = 74.0945$$

Selanjutnya dengan cara yang sama menghitung semua data yang ada hingga selesai. Setelah didapat jarak maka didapat nilai *centroid* baru dengan perhitungan :

$$C = \frac{Xi}{n} \quad (4)$$

Dengan Xi = jumlah total *cluster* ke i

n = jumlah anggota *cluter*

sehingga didapat hasil:

C1 = $\frac{37310,62319}{39} = 956.6826446$

C2 = $\frac{1301637}{1} = 1301637$

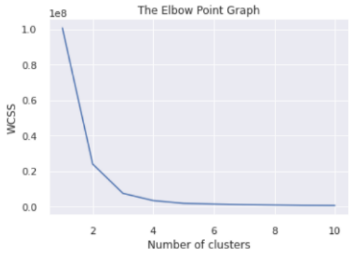
C3 = $\frac{606743.3}{1} = 606743.3$

C4 = $\frac{176651.7}{4} = 4462.9277$

C5 = $\frac{91605.06}{12} = 7633.755$

Hasil pemodelan dengan python menggunakan metode elbow untuk menentukkan jumlah cluster yang terbentuk. Gambar 5 adalah hasil dari pemodelan pada pemograman python.

```
[ ] #Plotting The Elbow graph
    plt.plot(range(1, 11), wcss)
    plt.title('The Elbow Point Graph')
    plt.xlabel('Number of clusters')
    plt.ylabel('WCSS')
    plt.show()
```



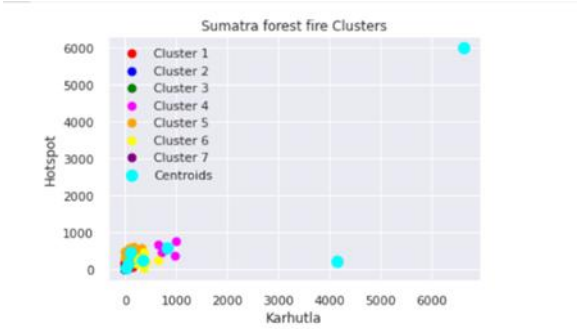
Gambar 5. Grafik Metode Elbow

4.4.4. **Mengelompokkan Objek**

Setelah didapat nilai centroid yang baru maka dapat dilakukan pengelompokkan data. Hasil yang didapat setelah perhitungan dan pengelompokkan melalui tool pemrograman python maka hasil pengelompokkannya dapat dilihat pada tabel 4. Berisi jumlah data , karakteristik, dan kriteria.

Tabel 4 Hasil Pengelompokkan

Cluster	Jumlah Data	Daerah Rawan	Karateristik	kriteria
1	39	Aceh Jaya, Bengkulu Selatan, muaro jambi, tulang bawang, Musirawas, Asahan, Nias	Luas karhutla 0-140 ha, jumlah hotspot 50-300	Rendah
2	1	Pesisir Selatan	Luas karhutla 6006 ha, jumlah hotspot 6006	Sangat tinggi
3	1	Lampung Timur	Luas karhutla 4048 ha, jumlah hotspot 204	Tinggi
4	4	Muko-muko, Rokan hilir, Pasaman Barat, Solok	Luas karhutla 600-1000 ha, jumlah hotspot 400-800	CukupTinggi
5	12	Bengkulu Utara, Kerinci, Lampung selatan, Lahat	Luas karhutla 0-100 ha, jumlah hotspot 300-650	Cenderung rendah
6	9	Aceh selatan, Dumai, Pasaman, Padang lawas	Luas karhutla 150-400 ha, jumlah hotspot 100-500	Sedang



Gambar 6. Grafik Pengelompokkan

Gambar 6 adalah hasil pengelompokkan yang didapat dengan menggunakan bahasa pemrograman python.

4.4. Evaluation

Untuk mengetahui seberapa baik pengelompokkan data percobaan dilakukan dengan metode pengujian nilai silhouette coefision. Berikut perhitungan dari silhouette coefision dengan menggunakan persamaan 1, 2, dan 3 dengan tabel 5.

Tabel 5 Kelompok Cluster 1

Kab/Kota	Luas Karhutla	Hotspot
Aceh Jaya	99	68
Kota Lhokseumawe	0	204
Kota Subulussalam	57	142
Naganraya	140	73
Bengkulu Selatan	8	112

Tahap pertama mencari nilai a(i) menggunakan persamaan 1 :
Cluster 1:
a(i) = 1/39 √((99-0)^2+(68-204)^2) + √((99-57)^2+(68-142)^2)
+ √((99-140)^2+(68-73)^2)+ √((99-8)^2+(68-112)^2)
+ √((….(i-j)^2)
= 4143.117

Langkah selanjutnya adalah menentukan nilai d(i) yang merupakan perhitungan cluster 1 dengan cluster lain . Perhitungan kali ini menggunakan data cluster 2 dan 7

Cluster 1 dengan cluster 2
d(i) = √((99-6024)^2+(68-6006)^2) + √((0-6024)^2+(204-6006))^4
+ √((57-6024)^2+(142-6006)^2) + √((…(i-j)^2)
= 343424.838

Perhitungan dilanjutkan sampai semua data. Langkah selanjutnya menghitung b(i) dengan persamaan 2.
b(i) = min (343424.838 ; 30516.11)
= 30516.11
Setelah didapat nilai a(i) dan b(i) langkah selanjutnya menghitung silhouette coefisients sesuai dengan persamaan 3.
S(i) = 1--(4143.117/30516.11)=0.864

Perhitungan nilai silhouette untuk cluster yang lain mengikuti langkah – langkah seperti diatas. Sehingga didapat hasil pehitungan pada tabel 6 :

Tabel 6 Hasil Perhitungan Rata – Rata Silhoutte Coefision

Cluster	*Silhoutte Coefision*
1	0.8642
2	0.9619
3	0.9339
4	0.7159

```
[ ] for n_clusters in range(2,7):
        clusterer = KMeans(n_clusters = n_clusters, max_iter=1000, random_state=10)
        cluster_label = clusterer.fit_predict(X)
        silhouette_avg = silhouette_score(X, cluster_label)
        print("cluster = ", n_clusters, "Nilai Rata-Rata Silhouette :", silhouette_avg )

    cluster =  2 Nilai Rata-Rata Silhouette : 0.9619513687930891
    cluster =  3 Nilai Rata-Rata Silhouette : 0.9339929366627789
    cluster =  4 Nilai Rata-Rata Silhouette : 0.7159105177463375
    cluster =  5 Nilai Rata-Rata Silhouette : 0.6596590370568316
    cluster =  6 Nilai Rata-Rata Silhouette : 0.6188276760643399
```

Gambar 7. Hasil Silhoutte Coefision

Gambar 7 merupakan nilai silhouette coefisient dengan pengujian pada pemrograman python.

`Berdasarkan hasil pengelompokan menggunakan metode K-Means, terbentuk beberapa klaster yang merepresentasikan tingkat kerawanan kebakaran hutan dan lahan dengan karakteristik yang berbeda-beda. Interpretasi singkat masing-masing klaster adalah sebagai berikut:

1. Cluster 1 (Rendah)

Cluster ini didominasi oleh daerah dengan luas kebakaran relatif kecil dan jumlah hotspot yang rendah hingga sedang. Wilayah pada cluster ini tergolong cukup aman, namun tetap memerlukan pemantauan karena masih ditemukan kemunculan hotspot yang berpotensi berkembang jika tidak ditangani sejak dini.

2. Cluster 2 (Sangat Tinggi)

Cluster ini hanya terdiri dari satu wilayah, namun memiliki luas karhutla dan jumlah hotspot yang sangat besar dibandingkan klaster lainnya. Hal ini menunjukkan adanya kejadian kebakaran hutan yang ekstrem dan masif. Wilayah dalam cluster ini menjadi prioritas utama dalam mitigasi dan penanganan kebakaran.

3. Cluster 3 (Tinggi)

Cluster ini menunjukkan wilayah dengan luas kebakaran yang besar, tetapi jumlah hotspot relatif lebih rendah dibandingkan cluster 2. Kondisi ini mengindikasikan bahwa kebakaran yang terjadi kemungkinan bersifat terpusat namun berdampak luas, sehingga tetap dikategorikan sebagai wilayah berisiko tinggi.

4. Cluster 4 (Cukup Tinggi)

Wilayah dalam cluster ini memiliki kombinasi luas karhutla dan jumlah hotspot pada tingkat menengah hingga tinggi. Daerah pada cluster ini berpotensi mengalami peningkatan risiko kebakaran apabila terjadi kondisi cuaca ekstrem seperti kemarau panjang.

5. Cluster 5 (Cenderung Rendah)

Cluster ini terdiri dari wilayah dengan luas kebakaran kecil, namun jumlah hotspot cukup banyak. Hal ini mengindikasikan adanya aktivitas kebakaran skala kecil yang sering terjadi, sehingga diperlukan upaya pencegahan agar tidak berkembang menjadi kebakaran yang lebih luas.

6. Cluster 6 (Sedang)

Cluster ini menunjukkan daerah dengan tingkat kebakaran dan hotspot yang seimbang pada level menengah. Wilayah ini memerlukan strategi pengendalian dan pengawasan rutin, karena berpotensi berpindah ke kategori risiko lebih tinggi.

7. Cluster 7 (Sangat Rendah)

Cluster ini didominasi oleh wilayah dengan luas kebakaran dan jumlah hotspot yang sangat rendah. Daerah pada cluster ini dapat dikategorikan sebagai wilayah relatif aman, namun tetap membutuhkan pemantauan sebagai langkah preventif.

5. Kesimpulan

Penelitian ini menggunakan 2 metode pengujian yaitu pengujian melalui tool Bahasa pemrograman python dan perhitungan menggunakan rumus. Namun kekurangan pada penelitian ini belum ada inisialisasi pada awal cluster.

Percobaan telah dilakukan dengan menggunakan perhitungan manual rumus dan pengujian lewat tool pada pemrograman python dapat disimpulkan bahwa kelompok daerah rawan bencana kebakaran hutan ada 7 cluster. Dengan perincian sebagai berikut, cluster 1 dengan status rendah, memiliki anggota sebanyak 39 data, cluster 2 status sangat tinggi memiliki 1 data, cluster 3 status tinggi memiliki 1 anggota data, cluster 4 status cukup tinggi memiliki 4 anggota data, cluster 5 status cenderung rendah memiliki 12 data, cluster 6 status sedang memiliki 9 data, dan cluster 7 dengan status sangat rendah memiliki 82 data.

Dari 7 cluster yang didapat, cluster 2 merupakan cluster yang memiliki nilai silhouette yang tertinggi sebesar 0.9619513687930891 dengan jumlah data uji 149 data. Analisa dari penelitian ini adalah kelompok atau cluster yang dihasilkan dengan metode k-mean memiliki nilai silhouette coefisient sebagai berikut cluster 1 sebesar 0.8642, cluster 2 sebesar 0.9619, cluster 3 sebesar 0.9339, cluster 4 sebesar 0.7159, cluster 5 sebesar 0.6596, cluster 6 sebesar 0.6188 dan cluster 7 sebesar 0.6007. Hasil evaluasi nilai silhouette dari semua cluster yaitu benilai positif dan mendekati satu sehingga dapat disimpulkan bahwa cluster yang terbentuk merupakan cluster yang baik dan anggota cluster sudah berada pada kluster yang tepat.

Dapat disimpulkan bahwa K-mean clustering dalam pemodelan pada penelitian ini mampu menghasilan cluster yang baik dengan dilihat dari nilai silhouette dari setiap cluster mendekati 1

Referensi

[1] R. Agung *et al.*, *Status Hutan dan Kehutanan Indonesia*. 2018.

[2] Supriyanto and dkk, "Analisis Kebijakan Pencegahan Dan Pengendalian Kebakaran Hutan Dan Lahan Di Provinsi Jambi," *Pembang. Berkelanjutan*, vol. 1, no. 1, pp. 94–104, 2018.

[3] A. Muzaki, R. Pratiwi, and S. R. Az Zahro, "Pengendalian Kebakaran Hutan Melalui Penguatan Peran Polisi Kehutanan Untuk Mewujudkan Sustainable Development Goals," *LITRA J. Huk. Lingkungan, Tata Ruang, dan Agrar.*, vol. 1, no. 1, pp. 22–44, 2021, doi: 10.23920/litra.v1i1.579.

[4] A. Yusuf, H. Hapsoh, S. H. Siregar, and D. R. Nurrochmat, "Analisis Kebakaran Hutan Dan Lahan Di Provinsi Riau," *Din. Lingkung. Indones.*, vol. 6, no. 2, p. 67, 2019, doi: 10.31258/dli.6.2.p.67-84.

[5] B. K. Amijaya, M. T. Furqon, and C. Dewi, "Clustering Titik Panas Bumi Menggunakan Algoritme Affinity Propagation," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 2, no. 10, pp. 3835–3842, 2018.

[6] F. Handayani, R. Teguh, and A. Lestari, "Pendeteksian Potensial Hotspot Data Satelit Check Algoritma Clustering K-Means," vol. 15, no. 2, pp. 164–173, 2021.

[7] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, "Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.

[8] R. T. Vulandari, W. L. Y. Saptomo, and D. W. Aditama, "Application of K-Means Clustering in Mapping of Central Java Crime Area," *Indones. J. Appl. Stat.*, vol. 3, no. 1, p. 38, 2020, doi: 10.13057/ijas.v3i1.40984.

[9] R. Gopikaramanan, T. Rameshkumar, B. Senthil Kumaran, and G. Ilangovan, *Novel control methodology for H-bridge cascaded multi level converter using predictive control methodology*, vol. 11, no. 5. 2015.

[10] K. P. Simanjuntak and U. Khaira, "Pengelompokkan Titik Api di Provinsi Jambi dengan Algoritma Agglomerative Hierarchical Clustering," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. April, pp. 7–16, 2021, [Online]. Available: https://journal.irpi.or.id/index.php/malcom/article/view/6

[11] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.

[12] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017, doi: 10.1109/EUMC.2008.4751704.

[13] Ahmad Harmain, P. Paiman, H. Kurniawan, K. Kusrini, and Dina Maulina, "Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas," *Tek. Teknol. Inf. dan Multimed.*, vol. 2, no. 2, pp. 83–89, 2022, doi: 10.46764/teknimedia.v2i2.49.

[14] E. Ikhsan, "Penerapan K-Means Clustering dari Log Data Moodle untuk Menentukan Perilaku Peserta pada Pembelajaran Daring," *Sistemasi*, vol. 10, no. 2, p. 414, 2021, doi: 10.32520/stmsi.v10i2.1285.

[15] Athifaturrofifah, R. Goejantoro, and D. Yuniarti, "Perbandingan Pengelompokan K-Means dan K-Medoids Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus : Data Titik Panas Di Indonesia Pada 28 April 2018)," *J. EKSPONENSIAL*, vol. 10, no. 2, pp. 143–152, 2019.

[16] A. A. D. Sulistyawati and M. Sadikin, "Penerapan Algoritma K-Medoids Untuk Menentukan Segmentasi Pelanggan," *Sistemasi*, vol. 10, no. 3, p. 516, 2021, doi: 10.32520/stmsi.v10i3.1332.

[17] E. P. Ariesanto Akhmad, "Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran," *J. Apl. Pelayaran dan Kepelabuhanan*, vol. 10, no. 2, p. 120, 2020, doi: 10.30649/japk.v10i2.83.

[18] N. Mirantika, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Nuansa Inform.*, vol. 15, no. 2, pp. 92–98, 2021, doi: 10.25134/nuansa.v15i2.4321.