

# Implementasi Metode CLARA Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)

Enni Lindrawati\*, Saiful Bahri, Umam Faqih Zubaedi, Vinnesa Patricia Carolina, Kusri, Dina Maulina

MTI Universitas Amikom Yogyakarta

Email: <sup>1,\*</sup>ennilindra@students.amikom.ac.id, <sup>2</sup>saifulbahri@students.amikom.ac.id, <sup>3</sup>faqihzubaedi99@students.amikom.ac.id, <sup>4</sup>vinnesa.p.c@students.amikom.ac.id, <sup>5</sup>kusrini@amikom.ac.id, <sup>6</sup>dina.m@amikom.ac.id

Submitted: 29/07/2022; Accepted: 30/08/2022; Published: 30/08/2022

**Abstrak**—Kebakaran hutan atau lahan merupakan peristiwa yang sering terjadi di berbagai negara di dunia yang membutuhkan penanganan serius dari semua pihak karena menimbulkan dampak bagi segala lini kehidupan. Oleh sebab itu dibutuhkan penanganan lebih dini, salah satunya dengan mengelompokkan wilayah-wilayah rawan kebakaran dengan menggunakan data titik panas. Titik panas dapat diperoleh dengan satelit dalam penelitian ini mengambil titik panas dari satelit NASA. Adapun data yang digunakan adalah latitude, longitude, brightness dan confidence. Metode yang digunakan adalah metode Clara Clustering karena metode ini mempunyai kelebihan tahan terhadap outlier/pencilan dan dapat digunakan dalam data jumlah besar. Penelitian ini menghasilkan kesimpulan bahwa metode Clara dapat memproses data titik panas sejumlah 16.579 dengan nilai *Silhouette Coefficient* terbaik sebesar 0.89 dengan 2 cluster sedangkan potensi yang dimiliki oleh cluster 1 termasuk kedalam potensi tinggi dengan hasil rata-rata *brightness* 367<sup>0</sup>K dan nilai *confidence* 1 atau high. Sedangkan cluster 2 memiliki potensi sedang dengan hasil rata-rata *brightness* 349<sup>0</sup>K dan nilai *confidence* 3 atau nominal.

**Kata Kunci:** Kebakaran Hutan; Metode Clara; *Silhouette Coefficient*; Titik Panas

**Abstract**—Forest or land fires are events that often occur in various countries in the world that require serious handling from all parties because they have an impact on all lines of life. Therefore, early treatment is needed, one of which is by grouping fire-prone areas using hotspot data. Hotspots can be obtained by satellite in this study taking hotspots from NASA satellites. The data used are latitude, longitude, brightness and confidence. The method used is the Clara Clustering method because this method has the advantage of being resistant to outliers and can be used in large amounts of data. This study concludes that the Clara method can process 16,579 hotspot data with the best *Silhouette Coefficient* value of 0.89 with 2 clusters while the potential possessed by cluster 1 is included in high potential with an average brightness of 3670K and a confidence value of 1 or high. Meanwhile, cluster 2 has moderate potential with an average brightness of 3490K and a confidence value of 3 or nominal.

**Keywords:** Forest Fires; Clara's Method; *Silhouette Coefficient*; Hotspot

## 1. PENDAHULUAN

Hutan merupakan tempat hidup berbagai organisme dengan keanekaragaman hayati, dan memiliki fungsi yang sangat baik dalam menjaga keseimbangan sumber daya alam. Bahan untuk kebutuhan hidup manusia [1].

Kebakaran sering terjadi baik di hutan maupun di darat. Hal ini dapat disebabkan oleh dua faktor utama: faktor alam dan manusia, faktor alam seperti layunya tanaman akibat kemarau panjang, dan aktivitas manusia seperti pembakaran terbuka ilegal untuk memperluas lahan [2]. Menurut Peraturan Menteri Kehutanan No. P.12/Menhut-II/2009, titik api adalah indikator kebakaran hutan yang mendeteksi tempat-tempat dengan suhu yang relatif lebih tinggi dari suhu lingkungan.

Titik panas merupakan salah satu sumber utama asap dan merupakan sumber pencemaran yang mengandung unsur-unsur yang berbahaya bagi kesehatan bila terhirup dalam konsentrasi tinggi [3]. Hotspot adalah suatu daerah yang suhu permukaannya relatif lebih tinggi dari daerah sekitarnya berdasarkan ambang batas suhu tertentu yang dipantau oleh satelit penginderaan jauh (bpbdbabelprov.go.id 2018).

Adapun parameter titik panas (hotspot) [4], yaitu 1) *Brightness* yaitu kecerahan *pixel* api yang diukur dalam satuan derajat Kelvin. 2) FRP (*fire radiative power*) yang menggambarkan *pixel* kekuatan radiasi api dalam MW (megawatt) diambil dari deteksi satelit. 3) *Confidence Hotspot* Tingkat kepercayaan mulai dari 0 sampai 100%.

Kebakaran hutan/lahan terdeteksi sebagai titik api dan muncul sebagai titik api. Hotspot adalah perbandingan titik panas yang relatif terhadap permukaan lainnya (Lapan, 2014). Asia Tenggara merupakan kawasan ditemukannya titik api, termasuk Indonesia (Tempo, 2015). Terdapat 259 titik api kebakaran hutan yang terekam pada tahun 2015 yang terdeteksi oleh kebakaran hutan satelit Terra/Aqua dengan keyakinan 80%. (Kompas, 2016). Efek kebakaran hutan mendorong pencegahan dini. Salah satu langkah pertama adalah menggunakan data hotspot untuk memprediksi area potensial kebakaran hutan. Hal ini dilakukan melalui proses pengelompokan sesuai informasi dalam data untuk mengetahui kemungkinan terjadinya kebakaran.

Teknik clustering merupakan teknik data mining yang tidak terawasi karena tidak menggunakan atribut untuk menyaring proses pembelajaran dan memperlakukan semua atribut secara setara.[5].

Penelitian yang berjudul “*Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)*” merupakan penelitian

sebelumnya dilakukan oleh Dyang Falila Pramesti dkk. Dalam penelitiannya data yang digunakan yaitu data *hotspot* dengan parameter *latitude*, *longitude*, *brightness*, *frp* (*fire radiativ power*), dan *confidence*, algoritma yang dipakai adalah K-Medoids. Hasil dari analisa didapat yakni hasil *clustering* dengan penggunaan 2 cluster dengan potensi yang dimiliki yaitu *cluster 1* menghasilkan rata-rata *brightness* sebesar  $344.47^0$  K dengan rata-rata *confidence* 87.08% sehingga termasuk dalam potensi tinggi dengan sedangkan *cluster 2* rata-rata *brightness* sebesar  $318.80^0$  K dan rata-rata *confidence* sebesar 58,73% sehingga termasuk dalam potensi sedang, [6].

Berdasarkan permasalahan paparan di atas maka kami melakukan penelitian yang dilakukan menjadi solusi agar mudah mengetahui potensi tingginya terjadi kebakaran hutan, yang merupakan pengembangan dari penelitian sebelumnya. Penelitian sebelumnya menggunakan K medoids sedangkan penelitian ini menggunakan clara karena clara merupakan algoritma untuk pengelompokan data dalam skala besar yang mengandung pencila menggunakan teknik pengambilan sample kemudian menerapkan algoritma PAM dengan perhitungan jarak antara objek terhadap pusat cluster dengan menggunakan jarak manhattan [7]. Untuk jumlah data penelitian dengan K-Medoids menggunakan jumlah data pengujian sebanyak 7352, penelitian ini menggunakan data uji sebanyak 16579 karena algoritma clara juga dapat digunakan untuk mengelompokkan data dengan jumlah yang besar[8]. Clara merupakan pengembangan dari PAM serta mengambil medoid dari sample [9].

## 2. METODOLOGI PENELITIAN

Adapun metode penelitian yang dilakukan sebagai berikut:

### 2.1 Pengumpulan Data

Adapun dataset yang digunakan adalah *hotspot* yang didapat dari situs resmi dari [kaggle.com](https://www.kaggle.com) yaitu Australian Bush fire satellite data (NASA). Adapun data yang diperoleh berupa nilai *brightness temperature*, *latitude*, *longitude* dan *confidence*. Dataset berupa data titik panas wilayah Australia tahun 2019 hingga 2020 berjumlah 105.714 baris, tetapi yang digunakan dalam proses pengujian sebanyak 16,579 baris.

### 2.2. Analisis Kebutuhan

Analisa kebutuhan dalam penelitian ini yaitu dengan menentukan variable-variable apa saja yang akan diimplementasikan pada metode CLARA untuk mengelompokkan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas (*hotspot*).

### 2.3 Proses Perhitungan

Clara digunakan dengan pendekatan sampling, setelah itu menerapkan algoritma PAM untuk memperoleh medoid yang optimal untuk sampel. Kualitas medoid yang dihasilkan diukur menggunakan rata-rata perbedaan jarak antara setiap objek di dataset dan medoid pada sampel. Sample diambil secara acak, medoid dari sampel diperkirakan mendekati nilai medoid dari dataset [10]. Tahapan analisis [11]:

1. Menetapkan jumlah kelompok k yang akan dibentuk dengan metode silhouette,
2. Melakukan pengelompokan data dengan algoritma CLARA:
3. Menampilkan hasil pengelompokan dengan algoritma CLARA
4. Perhitungan jarak dalam algoritma CLARA dengan jarak Manhattan sebagai berikut[7] :

$$d_{ij} = \sum_{k=1}^n |x_{ik} - x_{jk}| \quad (1)$$

Keterangan

$d_{ij}$  : jarak objek i ke objek j

$x_{ik}$  : objek i atribut ke- k

$x_{jk}$  : objek j atribut ke- k

5. Perhitungan *Silhouette Coefficient*

*Silhouette coefficient* terletak pada rentang nilai -1 sampai 1. Semakin dekat dengan nilai 1, maka semakin baik cluster hasil pengelompokan datanya. Sebaliknya jika nilai dekat dengan nilai -1, maka semakin buruk cluster hasil pengelompokannya [12]. Tahapan perhitungan *silhouette coefficient*[13]:

- a. Menghitung jarak rata-rata data dengan semua data yang berada dalam suatu cluster dengan persamaan berikut :

$$\alpha(i) = \frac{1}{|A|} \sum_{j \in A, j \neq i} d(i, j) \quad (2)$$

- b. Menghitung rata-rata data dengan semua data yang terletak dalam *cluster* lain, selanjutnya mengambil nilai yang paling kecil dengan persamaan berikut :

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (3)$$

c. Menghitung nilai *silhouette coefficient* dengan persamaan berikut :

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (4)$$

## 2.4 Pengujian Sistem dan Analisis Hasil

Dalam tahap pengujian Sistem dan Analisis Hasil ini menggunakan pemrograman R studio dan R project dengan versi 4.2.1. Aplikasi R pada dasarnya berbasis teks atau berbasis baris perintah, mengharuskan pengguna untuk mengetik perintah tertentu dan mengingat perintah. Setidaknya saat menggunakan R untuk melakukan aktivitas analisis data harus selalu memiliki perintah yang siap digunakan [14].

## 2.5 Pengambilan Kesimpulan

Setelah melakukan tahap pengujian dan Analisis hasil maka diambil kesimpulan.

# 3. HASIL DAN PEMBAHASAN

## 3.1 Data

Contoh data yang digunakan dalam penelitian ini terdapat pada tabel 1.

**Table 1.** Data

no	lalitude	longitude	Bright_ti4	confidence
1	-3748861	14963156	341.1	n
2	-344611	15088142	328.5	l
3	-3394823	15121292	341.1	n
4	-3445618	15087723	328.5	n
5	-3160223	15015147	367	h
6	-2926867	14946602	367	h
7	-2926898	1494623	356.2	n
8	-3160248	15014816	355.2	n
9	-2926928	14945857	340.6	n
10	-2934156	14804869	367	n

## 3.2 Normalisasi data

Normalisasi data adalah teknik desain database yang membuat tabel dengan atribut yang tidak sesuai dengan kebutuhan sistem[15]. Data sebelum dilakukan perhitungan harus dinormalisikan terlebih dahulu. Yaitu dengan mengubah type data. Type data semula adalah char diubah menjadi numerik (integer). Untuk nilai confidence yang semula berupa kategori “n”, “l”, dan “h” divectorkan menjadi “3”, “2” dan “1”.

**Table 2.** Data Hasil Normalisasi

no	lalitude	longitude	Bright_ti4	confidence
1	-3748861	14963156	341	3
2	-344611	15088142	328	2
3	-3394823	15121292	341	3
4	-3445618	15087723	328	3
5	-3160223	15015147	367	1
6	-2926867	14946602	367	1
7	-2926898	1494623	356	3
8	-3160248	15014816	355	3
9	-2926928	14945857	340	3
10	-2934156	14804869	367	1

## 3.3 Pengujian Cluster

Sejumlah uji klaster (k) dan uji volume data dilakukan. Tes dijalankan 5 kali dan nilai rata-rata digunakan sebagai hasil tes akhir.

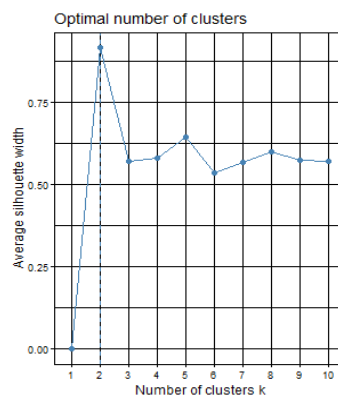
### 3.3.1 Pengujian Pengaruh Jumlah cluster (k)

Pengujian dampak jumlah cluster dilakukan dengan menggunakan 1658 data uji dan mempertimbangkan nilai *Silhouette Coefficient* yang dihasilkan untuk menentukan jumlah cluster yang optimal. Nilai cluster (k) yang akan diuji adalah nilai k antara 2 dan 10. Tabel 3 menunjukkan hasil pengujian.

**Table 3.** Hasil Pengujian Jumlah cluster

Jumlah Data Uji	Nilai K	Rata-Rata Nilai Silhouette Coefficient yang dihasilkan
1658	2	0,87
	3	0,56
	4	0,57
	5	0,62
	6	0,53
	7	0,55
	8	0,58
	9	0,55
	10	0,54

Proses pengujian menghasilkan kualitas *cluster* terbaik dengan penggunaan nilai k=2 dan nilai rata-rata *Silhouette Coefficient* paling tinggi adalah 0.87. Grafik ditunjukkan pada Gambar 2.



**Gambar 2.** Rata-rata *Silhouette Coefficient*

Grafik memperlihatkan semakin menggunakan data yang banyak dengan nilai k, maka akan memperoleh nilai *Silhouette Coefficient* relatif semakin rendah. Hal ini disebabkan oleh semakin banyaknya tetangga terdekat yang digunakan dalam proses *clustering*. Dalam hal ini, jarak ketetangaan diperhitungkan dan dimasukkan dalam proses perhitungan. Oleh karena itu, jarak antar objek dalam suatu *cluster* bertambah dan jarak antara *cluster* yang satu dengan *cluster* lainnya berkurang.

### 3.3.2 Pengujian Pengaruh Jumlah Data

Pengujian ini dilakukan untuk mengetahui pengaruh banyaknya data terhadap hasil dari pengujian sebelumnya. pengujian ini menggunakan data sebesar 10%, 25%, 50%, 75% dan 100% dari data sampel yang digunakan pada penelitian ini. Hasil pengujian ditunjukkan pada Tabel 4.

**Table 4.** Data Hasil Pengujian dengan Jumlah data

Jumlah Cluster	Jumlah Data	Rata-Rata Nilai Silhouette Coefficient
2	1658	0.87
	4145	0.88
	8290	0.87
	12434	0.87
	16579	0.89

Pada hasil proses pengujian, jumlah data terbaik diperoleh pada penggunaan 16.579 data dengan nilai *Silhouette Coefficient* tertinggi 0.89.

### 3.4 Analisis Penentuan Potensi Kebakaran Hutan/Lahan

**Table 5.** Hasil Analisis Penentuan Potensi

Cluster	Rata-rata Brightness ( $^{\circ}$ K)	Rata-rata Confidence
1	367	1
2	349	3

Dari tabel 5 tersebut dapat dianalisa bahwa potensi kebakaran hutan/lahan pada cluster 1 adalah tinggi karena rata-rata brightness  $367^0\text{K}$  dan rata-rata confidence 1 (high). Sedangkan cluster 2 memiliki potensi kebakaran sedang dengan rata-rata brightness  $349^0\text{K}$  dan confidence 3 (nominal).

#### 4. KESIMPULAN

Dapat menggunakan algoritma Clara untuk menangani clustering data potensi kebakaran hutan/lahan berdasarkan hotspot sesuai tahap perhitungan Clara. Kualitas clustering yang dihasilkan didasarkan pada proses hasil pengujian, diperoleh jumlah cluster 2 dan nilai koefisien Shihhoutte sebesar 0,89 pada dataset sebesar 16.579. Berdasarkan hasil tersebut, jumlah cluster dan jumlah data mempengaruhi kualitas hasil cluster. Analisis potensi kebakaran hutan/lahan di Klaster 1 lebih tinggi karena kecerahan rata-rata  $367^0\text{K}$  dan kepercayaan rata-rata 1 (tinggi). Cluster 2, di sisi lain, memiliki potensi pembakaran sedang, kecerahan rata-rata  $349^0\text{K}$ , dan keandalan 3 (nominal).

#### REFERENCES

- [1] V. Mala, A. Kusuma, M. T. Furqon, and L. Muflikhah, "Implementasi Metode Fuzzy Subtractive Clustering Untuk Pengelompokan," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 876–884, 2017.
- [2] A. Humam, M. Hidayat, A. Nurrochman, A. I. Anestatia, A. Yuliantina, and S. P. Aji, "Identifikasi Daerah Kerawanan Kebakaran Hutan dan Lahan Menggunakan Sistem Informasi Geografis dan Penginderaan Jauh di Kawasan Tanjung Jabung Barat Provinsi Jambi," *J. Geosains dan Remote Sens.*, vol. 1, no. 1, pp. 32–42, 2020, doi: 10.23960/jgrs.2020.v1i1.14.
- [3] F. Rozi, A. A. Akbar, and U. Kadaria, "Hubungan Sebaran Titik Panas (Hotspot) Terhadap Kesehatan Masyarakat Kota Pontianak," *J. Tek. Sipil*, vol. 20, no. 2, 2020.
- [4] 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.
- [5] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [6] 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.
- [7] P. R. Fitrayana and D. R. S. Saputro, "Algoritme Clustering Large Application (CLARA) untuk Menangani Data Outlier," *Prism. Pros. Semin. Nas. Mat.*, vol. 5, pp. 721–725, 2022, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [8] W. AS, M. K. Aidid, and M. Nusrang, "Pengelompokan Kabupaten/Kota Provinsi Sulawesi Selatan dan Barat Berdasarkan Angka Partisipasi Pendidikan SMA/SMK/MA Menggunakan K-Medoid dan CLARA," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 1, no. 3, p. 48, 2019, doi: 10.35580/variasiunm12899.
- [9] N. Thamrin and A. W. Wijayanto, "Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island," *Indones. J. Stat. Its Appl.*, vol. 5, no. 1, pp. 141–160, 2021, doi: 10.29244/ijsa.v5i1p141-160.
- [10] B. E. Susilowati and P. R. Sihombing, "Metode ROBPCA (Robust Principal Component Analysis) dan Clara (Clustering Large Area) pada Data dengan Outlier," *J. Ilmu Komput.*, vol. 13, no. 2, p. 11, 2020, doi: 10.24843/jik.2020.v13.i02.p04.
- [11] I. H. Rifa, H. Pratiwi, and Respatiwan, "Implementasi algoritma clara untuk data gempa bumi di indonesia," *Semin. Nas. Penelit. Pendidik. Mat.*, no. 2006, pp. 161–166, 2019.
- [12] P. Studi, T. Informatika, F. I. Komputer, U. Singaperbangsa, and A. K-means, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model," *J. Teknol. dan Indormasi*, vol. 12, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.
- [13] J. S. M. S. Mikroskil, "JSM STMIK Mikroskil 73," *Perbandingan Metod. Clust. Menggunakan Metod. Single Link. dan K-Means Pada Pengelompokan Dok.*, vol. Vol. 15, pp. 1–10, 2016.
- [14] M. Rosidi, "Metode Numerik Menggunakan R Untuk Teknik Lingkungan," p. 300, 2019, [Online]. Available: [https://bookdown.org/moh\\_rosidi2610/Metode\\_Numerik/](https://bookdown.org/moh_rosidi2610/Metode_Numerik/)
- [15] S. Mulyati, B. A. Sujatmoko, T. I. M. Wira, R. Afif, and R. A. Pratama, "Normalisasi Database Dan Migrasi Database Untuk Memudahkan Manajemen Data," *Sebatik*, vol. 22, no. 2, pp. 124–129, 2018, doi: 10.46984/sebatik.v22i2.319.