

**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)**

**Comparison of K-Means And K-Medoids Grouping Data on Potential Forest / Land Fires
Based on Hotspots Distribution
(Case study : Hotspots Data in Indonesia on April 28, 2018)**

Athifaturrofifah¹, Rito Goejantoro², dan Desi Yuniarti³

^{1,2}Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

³Laboratorium Statistika Ekonomi dan Bisnis FMIPA Universitas Mulawarman

E-mail: athifah.rofifah@gmail.com

Abstract

The cases of forest/land fires in Indonesia seem endless, almost every year in the dry season similar problems always occur. Some areas in Indonesia often occur in forest fires and result in losses of up to trillions of rupiah. Various ways have been made to help the government in minimizing the potential for forest or land fires, one of them is by monitoring hot spots. In this study using data hot spots with parameters of latitude, longitude, brightness, fire radiation power and confidence by using the method of grouping K-Means and K-Medoids. The difference between these two methods is that the K-means method uses the mean as the center of the cluster, while K-Medoids uses representative objects (medoids) as the center of the cluster. This study aims to compare the results of the grouping of K-Means method with K-Medoids by using 42 data. The results of this study indicate that the K-Means method produces Silhouette Coefficient scores greater than K-Medoids. So that, K-Means can provide more accurate grouping results with a greater Silhouette Coefficient value.

Keywords : K-Medoids, K-Means, Silhouette Coefficient, hot spots

Pendahuluan

Analisis pengelompokan adalah sebuah proses dalam pengelompokan tertentu terhadap suatu kumpulan objek ke dalam kelas-kelas yang mirip. Sebuah *cluster* adalah kumpulan dari objek yang mirip dengan satu sama lain pada *cluster* yang sama dan memiliki ketidakmiripan dengan objek di *cluster* yang lain. Analisis pengelompokan atau analisis *cluster* yang merupakan salah satu fungsi data *mining*, dapat digunakan sebagai alat untuk mengelompokkan data potensi kebakaran hutan (Miller & Han, 2009).

Dampak kebakaran hutan/lahan mendorong berbagai pihak untuk melakukan pencegahan secara dini. Mengingat bencana kebakaran hutan/lahan dapat diperparah saat memasuki musim-musim kemarau. Salah satu langkah awal yang dapat dilakukan dengan memperkirakan wilayah-wilayah yang berpotensi terhadap peristiwa kebakaran hutan/lahan. Proses pengelompokan dilakukan dengan menggunakan analisis pengelompokan, yaitu metode *K-Means* dan *K-Medoids* (Pramessti, 2017).

Metode *K-Means* melakukan pengelompokan data atau objek ke dalam *K* buah kelompok atau *cluster*. Pada setiap *cluster* terdapat titik pusat (*centroid*) yang merepresentasikan *cluster* tersebut. Sedangkan pengelompokan metode *K-Medoids* menggunakan objek representatif yang disebut *medoids* sebagai titik pusatnya (Suyanto, 2017).

Pada penelitian sebelumnya, telah dilakukan Pramessti (2017) terhadap proses pengelompokan pada kasus pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas di Wilayah Asia Tenggara pada tahun 2015. Penelitian yang dilakukan menggunakan metode *K-Medoids*.

Berdasarkan permasalahan diatas, penulis tertarik untuk membandingkan pengelompokan *K-Means* dan *K-Medoids* pada pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas di Indonesia pada tahun 2018.

Data Mining

Data mining adalah langkah analisis terhadap proses penemuan pengetahuan di dalam basis data atau *knowledge discovery in databases* yang disingkat KDD. Pengetahuan yang di maksud dapat berupa pola data atau relasi antar data (yang tidak diketahui sebelumnya)

Secara umum kegunaan *data mining* dapat dibagi menjadi dua yaitu deskriptif dan prediktif. Deskriptif berarti *data mining* digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data. Sedangkan prediktif berarti *data mining* digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi.

Berdasarkan fungsionalitasnya, tugas-tugas *data mining* bisa dikelompokkan ke dalam enam

kelompok diantaranya yaitu pengklasifikasian, pengelompokan, regresi, deteksi anomali, analisis asosiasi, dan perangkuman (Suyanto, 2017).

Standarisasi data

Sebelum melakukan proses analisis, salah satu yang perlu diperhatikan adalah apakah satuan data mempunyai perbedaan yang besar. Sebagai contoh, jika variabel penghasilan mempunyai satuan juta (0000.0000), sedangkan usia seseorang hanya mempunyai satuan puluhan (00), maka perbedaan yang mencolok ini akan membuat perhitungan jarak (*distance*) menjadi tidak valid (Santoso, 2015).

Untuk mengatasi masalah tersebut dapat digunakan teknik standarisasi dengan menentukan nilai rata-rata dan varian dalam persamaan berikut (Prasetyo, 2014):

$$\bar{x}_j = \frac{1}{N} \times \sum_{i=1}^N x_{ij} \tag{1}$$

$$\sigma_j^2 = \frac{1}{N-1} \times \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 \tag{2}$$

$$\hat{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \tag{3}$$

Dimana:

N = banyak data

x_{ij} = data ke- i pada variabel ke- j dengan

$i=1,2,\dots,N$ dan $j= 1,2,\dots,n$

\bar{x}_j = rata-rata pada variabel ke- j

s_j^2 = varian

s_j = standar deviasi

\hat{x}_{ij} = standarisasi data ke- i variabel ke- j

Metode K-Means

Menurut Prasetyo (2014), metode *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi *set* data ke dalam sejumlah K *cluster* yang sudah ditetapkan di awal. Dalam metode *K-Means* setiap *cluster* dari K *cluster* diwakili oleh titik tunggal yang disebut *centroid* c .

Meskipun konsep yang dimaksud untuk data-data yang berkumpul dalam satu *cluster* adalah data-data yang mirip, tetapi kuantitas yang digunakan untuk mengukurnya adalah ketidakmiripan (*dissimilarity*). Ukuran yang digunakan untuk ketidakmiripan dalam penelitian ini adalah jarak *Euclidean*. Ketidakmiripan data terhadap setiap *centroid* dihitung dengan menggunakan persamaan berikut.

$$d(x_{ij}, c_{Kj}) = \sqrt{\sum_{j=1}^N (x_{ij} - c_{Kj})^2} \tag{4}$$

Dimana $d(x_{ij}, c_{Kj})$ menyatakan ketidakmiripan dari data ke- i untuk variabel ke- j ke masing-masing *centroid* ke- K pada variabel ke- j .

Selanjutnya dipilih ketidakmiripan yang paling kecil sebagai *cluster* yang akan diikuti sebagai relokasi data pada *cluster* di sebuah iterasi. Relokasi sebuah data dalam *cluster* yang diikuti, dapat dinyatakan dengan nilai keanggotaan a yang bernilai 0 atau 1, di mana nilai 0 jika tidak menjadi anggota sebuah *cluster* dan 1 jika menjadi anggota sebuah *cluster*, seperti dinyatakan oleh persamaan berikut:

$$a_{iK} = \begin{cases} 1 & \text{min } d(x_{ij}, c_{Kj}) \\ 0 & \text{lainnya} \end{cases} \tag{5}$$

Sementara untuk mendapatkan titik *centroid* c_{Kj} didapatkan dengan menghitung rata-rata setiap fitur dari semua data yang tergabung dalam setiap *cluster* dinyatakan oleh persamaan berikut:

$$c_{Kj} = \frac{1}{n_K} \sum_{i=1}^{n_K} x_i \tag{6}$$

Dimana c_{Kj} adalah titik *centroid* ke- K pada variabel ke- j dimana $K = 1,2,\dots,p$, $j = 1,2,\dots,n$ dan n_K adalah jumlah data yang tergabung dalam sebuah *cluster*, sedangkan x_{iK} adalah data ke- i yang bergabung dalam *cluster* ke- K .

Metode dari *K-Means* adalah sebagai berikut:

1. Memilih objek K sebagai pusat *cluster* awal
2. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *centroid*
3. Menetapkan masing-masing objek ke *cluster* yang objeknya paling mirip, berdasarkan jarak antara objek terhadap setiap *centroid*
4. Memperbarui *centroid* yaitu menghitung nilai *mean* dari objek untuk setiap *cluster*
5. Mengulangi langkah 2-4 sampai tidak ada lagi perubahan *centroid*
6. Analisis selesai jika sudah tidak terdapat perubahan *centroid* atau pusat *cluster*

Metode K-Medoids

Metode *K-Medoids* merupakan metode yang menggunakan objek representatif yang disebut *medoids* sebagai titik pusat atau *centroid*. Metode *K-Medoids* melakukan partisi dengan cara meminimalkan jumlah *dissimilarity* antara setiap objek i dan objek representatif terdekat. Setiap objek yang tersisa dikelompokkan dengan objek representatif yang paling mirip dan perhitungan jarak dihitung dari jarak antar masing-masing data (Suyanto, 2017).

Pada metode sebelumnya yaitu *K-Means*, dikatakan sensitif terhadap *outlier* karena objek dengan beberapa nilai yang sangat besar dapat secara suportif mengubah distribusi data. Alih-alih mengambil nilai rata-rata objek-objek di dalam sebuah *cluster* sebagai titik acuan, seseorang dapat mengambil objek representatif dalam sebuah *cluster*, yang disebut *medoids*, yang merupakan titik terpusat dalam sebuah *cluster*. Dengan demikian metode partisi masih dapat dilakukan

berdasarkan prinsip meminimalkan jumlah ketidaksamaan antara masing-masing objek dan dengan titik acuan yang sesuai yang menjadi dasar metode *K-Medoids* (Suyanto, 2017).

Partitioning Around Method (PAM) adalah salah satu metode *clustering* tipe *K-Medoids*. Metode ini menentukan *K cluster* pada *n* objek dengan terlebih dahulu menentukan objek representatif (*medoids*) untuk setiap *cluster*. Set awal *medoids* dapat dipilih secara acak. Kemudian secara iteratif mengganti salah satu *non-medoids* selama jarak total pengelompokan yang dihasilkan ditingkatkan (Suyanto, 2017).

PAM mencoba untuk menentukan partisi *K* untuk *n* objek. Setelah pemilihan awal *K medoids* metode tersebut berulang kali mencoba untuk membuat pilihan yang lebih baik dari *medoids* dengan menganalisis semua kemungkinan pasangan objek sedemikian rupa sehingga satu objek adalah *medoids* dan yang lainnya tidak. Ukuran kualitas pengelompokan terbaik dihitung untuk setiap kombinasi tersebut, pilihan terbaik dari titik dalam satu iterasi dipilih sebagai *medoids* untuk iterasi berikutnya (Suyanto, 2017).

Metode *K-Medoids* adalah sebagai berikut (Han & Kamber, 2006):

1. Memilih *K* objek secara acak pada sekumpulan *N* objek sebagai objek representatif o_m (*medoids*)
2. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *medoids* seperti dinyatakan oleh persamaan berikut:

$$d(x_{ij}, o_{mj}) = \sqrt{\sum_{j=1}^N (x_{ij} - o_{mj})^2} \quad (7)$$

Dimana $d(x_{ij}, o_{mj})$ adalah jarak dari data ke-*i* pada variabel ke-*j* terhadap *medoids* ke-*m* pada variabel ke-*j* dimana $m = 1, 2, \dots, K$ dan $j = 1, 2, \dots, N$

3. Menetapkan setiap objek ke gugus yang sesuai dengan *medoids* terdekat dan menghitung fungsi obyektif yang merupakan jumlah disimilaritas dari semua objek ke *non medoids* terdekat berdasarkan jarak antara objek terhadap setiap *medoids* yang paling minimum
4. Memilih secara acak objek yang tidak representatif o_h (*non medoids*)
5. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *non medoids* seperti dinyatakan oleh persamaan berikut:

$$d(x_{ij}, o_{hj}) = \sqrt{\sum_{j=1}^N (x_{ij} - o_{hj})^2} \quad (8)$$

Dimana $d(x_{ij}, o_{hj})$ adalah jarak dari data ke-*i* pada variabel ke-*j* terhadap *non medoids*

ke-*h* pada variabel ke-*j* dimana $h = 1, 2, \dots, n-K$ dan $j = 1, 2, \dots, N$

6. Menetapkan setiap objek ke gugus yang sesuai dengan *non medoids* terdekat dan menghitung fungsi obyektif yang merupakan jumlah disimilaritas dari semua objek ke *non medoids* terdekat berdasarkan jarak antara objek terhadap setiap *medoids* yang paling minimum
7. Menghitung selisih dari fungsi obyektif
8. Mengganti *medoids* dengan *non medoids* apabila pertukaran semacam itu mengurangi fungsi obyektif
9. Mengulangi langkah 4-8 sampai tidak ada perubahan objek representatif
10. Analisis selesai jika sudah tidak terdapat perubahan objek representatif

Validasi Data Hasil Pengelompokan

Menurut Kaufman dan Rousseeuw (1990), salah satu metode evaluasi yang dapat digunakan untuk melihat kualitas dan kekuatan *cluster* adalah metode *Silhouette Coefficient*. Metode ini merupakan metode validasi *cluster* yang menggabungkan metode *cohesion* dan *separation*. Tahapan perhitungan *Silhouette Coefficient* adalah sebagai berikut

1. Menghitung rata-rata jarak dari suatu data ke-*i* dengan semua data yang berada pada satu *cluster* yang sama dengan menggunakan persamaan (9).

$$a_i = \frac{1}{n_K - 1} \sum_{r=1}^{n_K-1} d_{i,r}, r \neq 1 \quad (9)$$

dimana $K = 1, 2, \dots, p$.

2. Menghitung rata-rata jarak suatu data ke-*i* dengan semua data yang berada pada *cluster* yang berbeda dengan menggunakan Persamaan (11) dan diambil nilai terkecilnya.

$$b_i = \min d_i(K), r \neq 1 \quad (10)$$

dengan rumus jarak suatu data ke-*i* dengan semua data pada *cluster* yang berbeda adalah

$$d_i(K) = \frac{1}{n_K} \sum_{r=1}^{n_K} d_{i,r} \quad (11)$$

Dimana $K = 1, 2, \dots, p$.

3. Menghitung nilai *Silhouette Coefficient* untuk setiap data ke-*i*

$$SC_1(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}}, i = 1, 2, \dots, N \quad (12)$$

Nilai *SC* dari sebuah *cluster* ($SC_2(K)$) didapatkan dengan menghitung rata-rata nilai $SC_1(i)$ semua data yang bergabung dalam *cluster* tersebut dengan menggunakan persamaan berikut.

$$SC_2(K) = \frac{1}{n_K} \sum_{x_i \in C_k} SC_1(i) \tag{13}$$

Setelah itu nilai SC global didapatkan dengan menghitung rata-rata nilai $SC_2(K)$ dari semua $cluster$ dengan menggunakan persamaan berikut

$$SC = \frac{\sum_{i=1}^p n_K \times SC_2(K)}{\sum_{i=1}^p n_K} \tag{14}$$

Dimana:

- a_i : rata-rata jarak data ke- i dengan semua data pada satu $cluster$ yang sama
- b_i : rata-rata jarak data ke- i dengan semua data pada $cluster$ yang berbeda
- $SC_1(i)$: nilai *Silhouette Coefficient* pada data ke- i
- $SC_2(K)$: nilai *Silhouette Coefficient* pada $cluster$ ke- K
- SC : nilai *Silhouette Coefficient* global
- x_i : data pengamatan ke- i
- C_k : $cluster$ ke- K
- n_k : jumlah data dalam $cluster$ ke- K
- p : jumlah $cluster$

Nilai *Silhouette Coefficient* adalah sebagai berikut.

Tabel 1. Nilai *Silhouette Coefficient*

No	Rentang Nilai SC	Keterangan
1.	$0,7 < SC \leq 1$	<i>Strong Structure</i>
2.	$0,5 < SC \leq 0,7$	<i>Medium Structure</i>
3.	$0,25 < SC \leq 0,5$	<i>Weak Structure</i>
4.	$SC \leq 0,25$	<i>No Structure</i>

Titik Panas

Titik panas atau *Hotspots* secara definisi dapat diartikan sebagai daerah yang memiliki suhu permukaan relatif lebih tinggi dibandingkan daerah disekitarnya berdasarkan ambang batas suhu tertentu yang terpantau oleh satelit penginderaan jauh (LAPAN, 2018).

NASA (2018) mendefinisikan bahwa data titik panas memiliki atribut untuk mendeteksi adanya kebakaran hutan. Atribut pada sensor satelit MODIS yang digunakan dalam penelitian yaitu :

- a) *Latitude* merupakan garis lintang dimana garis ini merupakan garis horizontal yang melingkari bumi ditarik dari arah barat hingga ke timur.
- b) *Longitude* merupakan garis bujur dimana garis ini merupakan garis vertikal yang ditarik dari kutub utara hingga ke kutub selatan.
- c) *Brightness Temperature* merupakan ukuran deskriptif dari pancaran radiasi dalam bentuk suhu yang dipancarkan ke atas pada bagian atmosfer bumi. *Brightness Temperature* merupakan fitur dasar pada citra penginderaan

jarak jauh yang di deteksi dari lokasi tertentu dan diukur dalam satuan ukur kelvin.

- d) *Fire Radiative Power (FRP)* merupakan gambaran kekuatan radiasi piksel api yang kemudian diintegrasikan dalam *Mega Watt (MW)*. FRP memberikan informasi pada *output* radiasi panas dari api yang terdeteksi. Jumlah radiasi energi panas dibebaskan per satuan waktu (FRP) dimana diduga berkaitan dengan tingkat bahan bakar yang dikonsumsi.
- e) *Confidence* merupakan tingkat kepercayaan di mana *hotspot* yang dipantau oleh satelit merupakan kejadian kebakaran yang sebenarnya terjadi di lapangan. Semakin tinggi selang kepercayaan, maka semakin tinggi pula potensi bahwa *hotspot* benar-benar merupakan kebakaran hutan atau lahan yang terjadi. Berikut tingkatan *confidence* dalam menggambarkan informasi *hotspot* ditunjukkan pada Tabel 2.

Tabel 2. Makna Selang Kepercayaan Dalam Informasi Hotspot

Tingkat Kepercayaan	Kelas	Tindakan
$0\% \leq C < 30\%$	Rendah	Perlu diperhatikan
$30\% \leq C < 80\%$	Nominal	Waspada
$80\% \leq C \leq 100\%$	Tinggi	Segera penanggulangan

Kebakaran Hutan

Di negeri kita tercinta saat ini, telah banyak bencana alam yang datang silih berganti. Hal ini dikarenakan pembangunan berwawasan lingkungan hidup telah terabaikan. Setiap daerah berlomba dengan dalih meningkatkan hasil pendapatan asli daerahnya. Namun jalan yang ditempuh banyak daerah tidak lagi menggunakan konsep pembangunan berwawasan lingkungan hidup. Akibatnya dapat kita lihat bersama dimana banyak hutan yang rusak oleh pembalakan liar, pengambil alihan fungsi hutan untuk kepentingan ekonomi sesaat, bencana kabut asap, kebakaran hutan dan sebagainya (Transtoto, 2009).

Hasil Penelitian Dan Pembahasan

Standarisasi Data

Standarisasi data adalah proses untuk menstandarkan data dalam jangkauan yang sama. Dalam melakukan standarisasi data digunakan Persamaan (1), (2) dan (3) untuk 42 data titik panas.

Analisis Pengelompokan dengan K-Means

Proses pengelompokan menggunakan metode *K-Means* adalah sebagai berikut:

- 1. Memilih objek K sebagai pusat $cluster$ awal. Pada penelitian ini digunakan jumlah $cluster$ K sebanyak 2. Masing-masing pusat $cluster$ awal

- yang telah ditentukan secara acak yaitu untuk c_1 menggunakan data ke-27, dan c_2 menggunakan data ke-1.
- Menghitung jarak Euclid untuk setiap objek terhadap masing-masing pusat *cluster* awal dengan menggunakan persamaan (4).

Tabel 3. Data Hasil Perhitungan Jarak Terhadap Masing-Masing Pusat *Cluster* Awal

Data	Jarak ke pusat <i>cluster</i> awal	
	1	2
x_1	5,52	0
x_2	1,321	4,438
x_3	1,336	4,355
x_4	1,008	4,821
\vdots	\vdots	\vdots
x_{41}	1,532	4,367
x_{42}	2,885	4,098

- Menetapkan masing-masing objek ke *cluster* terdekat berdasarkan jarak antara objek terhadap setiap pusat *cluster* yang paling minimum, menggunakan persamaan (5).

Tabel 4. Penetapan Objek Terhadap Setiap *Cluster*

Data	<i>Cluster</i> (C)	
	1	2
x_1	0	1
x_2	1	0
x_3	1	0
x_4	1	0
\vdots	\vdots	\vdots
x_{41}	1	0
x_{42}	1	0

- Memperbarui pusat *cluster* atau *centroid* yaitu menghitung nilai rata-rata dari semua objek yang bergabung dalam setiap *cluster* dengan menggunakan Persamaan (6).

Tabel 5. Data Hasil Perhitungan *Centroid*

	X_1	X_2	X_3
c_1	-0,253	-0,235	-0,223
c_2	1,872	1,737	1,653

- Mengulangi langkah 2-4 hingga tidak ada lagi perubahan *centroid*. Banyak iterasi yang dilakukan yaitu 2 iterasi. Karena tidak ada lagi perubahan *centroid*, oleh karena itu proses pengelompokan dihentikan. Data hasil pengelompokan untuk 2 *cluster* dapat dilihat pada Tabel 6. Dari Tabel 6, dapat diketahui bahwa data x_1 dengan nilai *Latitude* -8,20 dan nilai *Longitude* 119,07 untuk wilayah NTB, dengan nilai *brightness temperature* (X_1) sebesar 328,8°K, nilai *confidence* (X_2) sebesar 83% dan nilai *fire radiative power* (X_3) sebesar 44,4 *Mega Watt* masuk dalam *cluster* kedua. Sedangkan data x_2 dengan nilai *Latitude* 0,70

dan nilai *Longitude* 102,13 untuk wilayah Riau, dengan nilai *brightness temperature* (X_1) sebesar 317°K, nilai *confidence* (X_2) sebesar 57% dan nilai *fire radiative power* (X_3) sebesar 9,5 *Mega Watt* masuk dalam *cluster* pertama. Demikian seterusnya untuk penafsiran objek lainnya.

Tabel 6. Data Hasil Pengelompokan *K-Means*

Data	Wilayah	X_1	X_2	X_3	C_1	C_2
x_1	NTB	328,8	83	44,4	0	1
x_2	Riau	317	57	9,5	1	0
x_3	Riau	316,8	57	10,6	1	0
x_4	Riau	315,5	55	6,8	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{41}	Riau	319,6	48	11,1	1	0
x_{42}	Riau	327,1	36	16,6	1	0

Hasil pemetaan wilayah pengelompokan *K-Means* juga dapat dilihat pada Gambar 1, dengan keterangan sebagai berikut:

- Berpotensi Sedang
- Berpotensi Tinggi



Gambar 1. Pengelompokan wilayah yang berpotensi kebakaran hutan/lahan di Indonesia Metode *K-Means*

Berdasarkan gambar diatas, maka dapat diinterpretasikan bahwa wilayah yang ditandai dengan warna abu-abu berarti termasuk wilayah-wilayah yang tergabung dalam *cluster* pertama, yaitu wilayah yang memiliki potensi sedang dan dibutuhkan tindakan kewaspadaan terhadap terjadinya kebakaran. Sedangkan untuk wilayah yang ditandai dengan warna hitam berarti termasuk wilayah-wilayah yang tergabung dalam *cluster* kedua, yaitu wilayah yang memiliki potensi tinggi, titik-titik tersebut menunjukkan kebakaran yang sebenarnya dan dibutuhkan penanggulangan segera.

Analisis Pengelompokan dengan *K-Medoids*

Proses pengelompokan menggunakan metode *K-Medoids* adalah sebagai berikut:

- Memilih secara acak objek K sebagai objek representatif awal atau *medoids*. Pada penelitian ini digunakan jumlah *cluster* (K) sebanyak 2. Masing-masing *medoids* o_j yang telah ditentukan secara acak yaitu untuk *medoids* 1 menggunakan data ke-27, dan *medoids* 2 menggunakan data ke-1.

- Menghitung jarak Euclid untuk setiap objek terhadap masing-masing objek *medoids* menggunakan persamaan (7).

Tabel 7. Data Hasil Perhitungan Jarak Terhadap Masing-Masing *Medoids* (o_1, o_2)

Data	$d(x_{ij}, o_{1j})$	$d(x_{ij}, o_{2j})$
x_1	5,52	0
x_2	1,321	4,438
x_3	1,336	4,355
x_4	1,008	4,821
\vdots	\vdots	\vdots
x_{41}	1,532	4,367
x_{42}	2,885	4,098

- Menetapkan setiap objek ke gugus yang sesuai dengan *medoids* terdekat, berdasarkan jarak antara objek terhadap setiap *medoids* yang paling minimum.

Tabel 8. Data Hasil Penetapan Masing-Masing Objek ke *Medoids* Terdekat

Data	C_1	C_2	Minimum
x_1	5,52	0	0
x_2	1,321	4,438	1,321
x_3	1,336	4,355	1,336
x_4	1,008	4,821	1,008
\vdots	\vdots	\vdots	\vdots
x_{41}	1,532	4,367	1,532
x_{42}	2,885	4,098	2,885
	Total jarak minimum		66,562

Dari Tabel 8 dapat diketahui bahwa jumlah disimilaritas adalah 66,562 yang disebut sebagai fungsi obyektif.

- Memilih secara acak *non medoids*. Diketahui jumlah data adalah sebanyak 42 data sehingga terdapat 40 objek yang menjadi *non medoids*. Dari 40 *non medoids* yang telah ditentukan, akan dipilih secara acak 2 kombinasi dari *non medoids-non medoids*, sehingga diperoleh sebanyak 780 kombinasi. Objek yang dipilih secara acak menjadi *non medoids* dapat dilihat seperti pada Tabel 9. Pada Tabel 9, dapat diketahui bahwa yang menjadi *non medoids* 1 adalah objek ke-6, yang menjadi *non medoids* 2 adalah objek ke-38 dan yang menjadi *non medoids* 3 adalah objek ke-16 demikian seterusnya untuk penafsiran objek lainnya.
- Menghitung jarak Euclid setiap objek terhadap masing-masing *non medoids* berdasarkan Persamaan (8). Masing-masing *non medoids* (o_h) yang telah dipilih secara acak yaitu, untuk o_1 menggunakan data ke-6, dan o_2 menggunakan data ke-38.

Tabel 9. Objek-Objek yang Menjadi *Non Medoids*

<i>Non Medoids</i>	Objek	<i>Non Medoids</i>	Objek
o_1	x_6	o_{21}	x_{29}
o_2	x_{38}	o_{22}	x_{32}
o_3	x_{16}	o_{23}	x_{36}
o_4	x_2	o_{24}	x_{41}
o_5	x_7	o_{25}	x_{24}
o_6	x_{11}	o_{26}	x_{33}
o_7	x_3	o_{27}	x_{42}
o_8	x_{19}	o_{28}	x_{35}
o_9	x_{23}	o_{29}	x_{17}
o_{10}	x_{14}	o_{30}	x_{15}
o_{11}	x_{26}	o_{31}	x_8
o_{12}	x_{31}	o_{32}	x_{10}
o_{13}	x_{21}	o_{33}	x_{12}
o_{14}	x_{34}	o_{34}	x_{30}
o_{15}	x_{40}	o_{35}	x_{37}
o_{16}	x_9	o_{36}	x_{39}
o_{17}	x_{18}	o_{37}	x_4
o_{18}	x_{25}	o_{38}	x_{13}
o_{19}	x_{20}	o_{39}	x_5
o_{20}	x_{22}	o_{40}	x_{28}

Tabel 10. Data Hasil Perhitungan Jarak Terhadap *Non Medoids* (o_1, o_2)

Data	$d(x_{ij}, o_{1j})$	$d(x_{ij}, o_{2j})$
x_1	5,534	4,511
x_2	1,599	0,345
x_3	1,591	0,361
x_4	1,33	0,431
\vdots	\vdots	\vdots
x_{41}	2,062	1,047
x_{42}	3,513	2,524

- Menetapkan setiap objek ke gugus yang sesuai dengan *non medoids* terdekat, berdasarkan jarak antara objek terhadap setiap *non medoids* yang paling minimum.

Tabel 11. Data Hasil Penetapan Masing-Masing Objek ke *Non Medoids* Terdekat

Data	C_1	C_2	Minimum
x_1	5,534	4,511	4,511
x_2	1,599	0,345	0,345
x_3	1,591	0,361	0,361
x_4	1,33	0,431	0,431
\vdots	\vdots	\vdots	\vdots
x_{41}	2,062	1,047	1,047
x_{42}	3,513	2,524	2,524
	Total jarak minimum		57,530

Dari Tabel 11 dapat diketahui bahwa jumlah disimilaritas adalah 57,530 yang disebut sebagai fungsi obyektif.

- Menghitung selisih dari fungsi obyektif, dengan cara mengurangi fungsi obyektif

non medoids dengan fungsi obyektif medoids.
 Selisih = fungsi obyektif non medoids - fungsi obyektif medoids
 Selisih = 57,530-66,562
 = -9,032

Karena nilai selisih yang didapatkan kurang dari 0, maka tukar nilai medoids (o_1, o_2) dengan non medoids (o_1, o_2) karena objek-objek non medoids lebih baik dari objek medoids.

8. Dilakukan perhitungan yang sama pada iterasi berikutnya dengan mengganti kedua non medoids yang lama dengan kedua non medoids yang baru, sampai tidak ada lagi perubahan objek-objek representatif. Hasil dari keseluruhan iterasi yang dilakukan adalah dari iterasi 1 hingga iterasi ke 780.

Karena tidak ada lagi perubahan objek representatif pada iterasi ke 780, maka proses pengelompokan dihentikan. Selanjutnya dapat diketahui bahwa yang menjadi objek representatif terbaik adalah non medoids (o_{20}, o_{26}) yang diperoleh pada iterasi ke 595 dengan total jarak minimum sebesar 44,656.

Data hasil pengelompokan K-Medoids untuk 2 cluster dapat dilihat pada Tabel 12.

Tabel 12. Data Hasil Pengelompokan K-Medoids

Data	Wilayah	X_1	X_2	X_3	C_1	C_2
x_1	NTB	328,8	83	44,4	0	1
x_2	Riau	317	57	9,5	1	0
x_3	Riau	316,8	57	10,6	1	0
x_4	Riau	315,5	55	6,8	1	0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
x_{41}	Riau	319,6	48	11,1	1	0
x_{42}	Riau	327,1	36	16,6	1	0

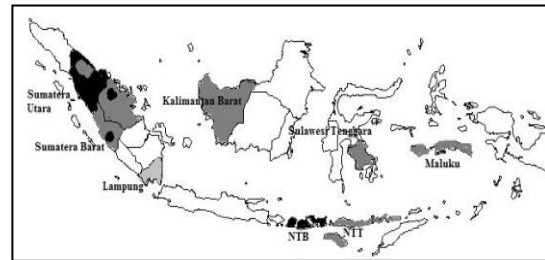
Sebagai contoh pada data x_1 dengan nilai Latitude -8,20 dan nilai Longitude 119,07 untuk wilayah NTB, dengan nilai brightness temperature (X_1) sebesar 328,8°K, nilai confidence (X_2) sebesar 83% dan nilai fire radiative power (X_3) sebesar 44,4 Mega Watt masuk dalam cluster kedua. Demikian seterusnya untuk penafsiran objek lainnya.

Hasil pemetaan wilayah pada pengelompokan ini juga dapat dilihat pada Gambar 2, dengan keterangan sebagai berikut:

- Berpotensi Sedang
- Berpotensi Tinggi

Berdasarkan Gambar 2, maka dapat diinterpretasikan bahwa daerah yang berwarna abu-abu termasuk wilayah yang tergabung dalam cluster pertama, yaitu wilayah yang memiliki potensi sedang artinya dibutuhkan tindakan kewaspadaan terhadap terjadinya kebakaran. Sedangkan untuk daerah berwarna hitam berarti termasuk wilayah yang tergabung dalam cluster kedua, yaitu wilayah yang memiliki potensi tinggi,

titik-titik tersebut menunjukkan kebakaran yang sebenarnya dan dibutuhkan penanggulangan segera.



Gambar 2. Pengelompokan wilayah yang berpotensi kebakaran hutan/lahan di Indonesia Metode K-Medoids

Validasi Hasil Pengelompokan dengan Metode K-Means

Hasil validasi cluster K-Means berdasarkan nilai SC setiap data ke-i sebagai berikut.

Tabel 13. Hasil Validasi Cluster Metode K-Means Berdasarkan Nilai SC untuk Setiap Data ke-i

No	Data	C	$SC_1(i)$	No	Data	C	$SC_1(i)$
1	x_1	2	0,393	22	x_{22}	2	0,308
2	x_2	1	0,679	23	x_{23}	1	0,651
3	x_3	1	0,674	24	x_{24}	1	0,507
4	x_4	1	0,696	25	x_{25}	1	0,628
5	x_5	1	0,67	26	x_{26}	1	0,191
6	x_6	1	0,597	27	x_{27}	1	0,669
7	x_7	1	0,666	28	x_{28}	1	0,395
8	x_8	1	0,685	29	x_{29}	1	0,103
9	x_9	1	0,569	30	x_{30}	1	0,565
10	x_{10}	1	0,71	31	x_{31}	1	0,575
11	x_{11}	1	0,531	32	x_{32}	1	0,545
12	x_{12}	1	0,619	33	x_{33}	1	0,698
13	x_{13}	1	0,665	34	x_{34}	1	0,711
14	x_{14}	1	0,642	35	x_{35}	1	0,634
15	x_{15}	1	0,678	36	x_{36}	1	0,671
16	x_{16}	2	0,399	37	x_{37}	1	0,565
17	x_{17}	1	0,519	38	x_{38}	1	0,653
18	x_{18}	2	0,253	39	x_{39}	1	0,496
19	x_{19}	2	0,248	40	x_{40}	1	0,633
20	x_{20}	1	0,662	41	x_{41}	1	0,666
21	x_{21}	1	0,685	42	x_{42}	1	0,352

Berdasarkan Tabel 13 dapat diketahui bahwa untuk data pertama masuk dalam cluster kedua, memiliki nilai SC sebesar 0,393 yang artinya data tersebut dapat dikatakan weak structure (lemah) untuk berada didalam cluster kedua. Data kedua masuk dalam cluster pertama, memiliki nilai SC sebesar 0,679 yang artinya data tersebut dapat dikatakan medium structure (baik) untuk berada di dalam cluster pertama. Demikian seterusnya untuk penafsiran objek lainnya.

Hasil validasi cluster berdasarkan nilai Silhouette Coefficient (SC) setiap cluster ke-K dapat dilihat pada Tabel 14.

Tabel 14. Hasil Validasi Cluster Metode K-Means Berdasarkan Nilai SC Setiap Cluster ke-K

Cluster (C)	SC(K)
1	0,591
2	0,320

Berdasarkan Tabel 14 dapat diketahui bahwa untuk *cluster* pertama memiliki nilai *SC* sebesar 0,591 yang artinya *cluster* tersebut dapat dikatakan *medium structure* (struktur *cluster* yang baik), yang artinya *cluster* tersebut dapat dikatakan *medium structure* (struktur *cluster* yang medium/baik), artinya pengelompokan objek *i* berada dalam *cluster* yang tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap *cluster*. Kumpulan objek *i* yang bergabung dalam *cluster* yang sama memiliki karakteristik yang mirip. *Cluster* kedua memiliki nilai *SC* sebesar 0,320 yang artinya *cluster* tersebut dapat dikatakan *weak structure* (struktur *cluster* yang lemah/buruk), artinya sedikit sekali objek *i* yang berada dalam *cluster* yang tepat dan sisanya bergabung pada *cluster* yang tidak semestinya. Menghitung nilai *SC* global yaitu dengan menghitung rata-rata nilai $SC_2(K)$ menggunakan Persamaan (13).

Hasil validasi untuk 2 *cluster* berdasarkan nilai *Silhouette Coefficient (SC)* global adalah sebesar 0,558 artinya pengelompokan secara keseluruhan dapat dikatakan *medium structure* (struktur *cluster* yang baik), artinya pengelompokan objek-objek berada dalam *cluster-cluster* yang tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap *cluster*. Kumpulan objek *i* yang bergabung dalam *cluster* yang sama memiliki karakteristik yang mirip.

Nilai-nilai $SC_1(i)$, $SC_2(K)$ dan *SC* inilah yang akan dijadikan acuan untuk membandingkan metode mana yang lebih baik. Semakin mendekati nilai 1 maka hasil pengelompokan yang dihasilkan tersebut semakin baik.

Validasi Hasil Pengelompokan dengan Metode K-Medoids

Hasil validasi *cluster K-Medoids* berdasarkan nilai *Silhouette Coefficient (SC)* setiap data ke-*i* dapat dilihat pada Tabel 15.

Berdasarkan Tabel 15 dapat diketahui bahwa untuk data pertama masuk dalam *cluster* kedua, memiliki nilai *SC* sebesar 0,377 yang artinya data tersebut dapat dikatakan *weak structure* (lemah) untuk berada didalam *cluster* kedua. Data kedua masuk dalam *cluster* pertama, memiliki nilai *SC* sebesar 0,647 yang artinya *cluster* tersebut dapat dikatakan *medium structure* (baik), untuk berada di dalam *cluster* pertama. Demikian seterusnya untuk penafsiran objek lainnya.

Hasil validasi *cluster* berdasarkan nilai *Silhouette Coefficient (SC)* setiap *cluster* ke-*K* dapat dilihat pada Tabel 16. Berdasarkan Tabel 16 dapat diketahui bahwa untuk *cluster* pertama memiliki nilai *SC* sebesar 0,569 yang artinya *cluster* tersebut dapat dikatakan *medium structure* (struktur *cluster* yang baik), yang artinya *cluster* tersebut dapat dikatakan *medium structure*

(struktur *cluster* yang medium/baik), artinya pengelompokan objek *i* berada dalam *cluster* yang tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap *cluster*. Kumpulan objek *i* yang bergabung dalam *cluster* yang sama memiliki karakteristik yang mirip. *Cluster* kedua memiliki nilai *SC* sebesar 0,286 yang artinya *cluster* tersebut dapat dikatakan *weak structure* (struktur *cluster* yang lemah/buruk), artinya sedikit sekali objek *i* yang berada dalam *cluster* yang tepat dan sisanya bergabung pada *cluster* yang tidak semestinya. Kumpulan objek *i* yang bergabung dalam *cluster* yang sama memiliki karakteristik yang berbeda.

Tabel 15. Hasil Validasi *Cluster* Metode *K-Medoids* Berdasarkan Nilai *SC* untuk Setiap Data ke-*i*

No	Data	C	SC ₁ (i)	No	Data	C	SC ₁ (i)
1	x ₁	2	0,377	22	x ₂₂	2	0,386
2	x ₂	1	0,647	23	x ₂₃	1	0,629
3	x ₃	1	0,64	24	x ₂₄	1	0,436
4	x ₄	1	0,671	25	x ₂₅	1	0,588
5	x ₅	1	0,66	26	x ₂₆	1	0,171
6	x ₆	1	0,57	27	x ₂₇	1	0,651
7	x ₇	1	0,651	28	x ₂₈	1	0,289
8	x ₈	1	0,658	29	x ₂₉	2	-0,1
9	x ₉	1	0,52	30	x ₃₀	1	0,523
10	x ₁₀	1	0,693	31	x ₃₁	1	0,518
11	x ₁₁	1	0,479	32	x ₃₂	1	0,529
12	x ₁₂	1	0,581	33	x ₃₃	1	0,584
13	x ₁₃	1	0,648	34	x ₃₄	1	0,693
14	x ₁₄	1	0,607	35	x ₃₅	1	0,612
15	x ₁₅	1	0,655	36	x ₃₆	1	0,656
16	x ₁₆	2	0,366	37	x ₃₇	1	0,511
17	x ₁₇	1	0,493	38	x ₃₈	1	0,616
18	x ₁₈	2	0,347	39	x ₃₉	1	0,443
19	x ₁₉	2	0,342	40	x ₄₀	1	0,608
20	x ₂₀	1	0,633	41	x ₄₁	1	0,639
21	x ₂₁	1	0,671	42	x ₄₂	1	0,323

Tabel 16 Hasil Validasi *Cluster* Metode *K-Medoids* Berdasarkan Nilai *SC* Setiap *Cluster* ke-*K*

<i>Cluster (C)</i>	<i>SC(K)</i>
1	0,569
2	0,286

Menghitung nilai *SC* global yaitu dengan menghitung rata-rata nilai $SC_2(K)$ dengan menggunakan Persamaan (13).

Hasil validasi untuk 2 *cluster* berdasarkan nilai *Silhouette Coefficient (SC)* global adalah sebesar 0,529 artinya pengelompokan secara keseluruhan dapat dikatakan *medium structure* (struktur *cluster* yang baik), artinya pengelompokan objek-objek berada dalam *cluster-cluster* yang tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap *cluster*. Kumpulan objek *i* yang bergabung dalam *cluster* yang sama memiliki karakteristik yang mirip.

Nilai-nilai $SC_1(i)$, $SC_2(K)$ dan *SC* inilah yang akan dijadikan acuan untuk memvalidasi kedua metode pengelompokan. Semakin mendekati nilai

1 maka hasil pengelompokan yang dihasilkan tersebut semakin baik.

Perbandingan Hasil Pengelompokan

Pada bagian ini dilakukan perbandingan dari kedua metode berdasarkan nilai *SC* global. Perbandingan hasil validasi *cluster* dari kedua metode dapat sebagai berikut.

Tabel 17. Perbandingan Hasil Validasi *Cluster* Kedua Metode Berdasarkan Nilai *SC* Global

<i>Metode</i>	<i>SC</i>
<i>K-Means</i>	0,558
<i>K-Medoids</i>	0,529

Dari Tabel 17 dapat diketahui bahwa Nilai *SC* untuk validasi data hasil pengelompokan data potensi kebakaran hutan/lahan berdasarkan persebaran titik panas di Indonesia dengan metode *K-Means* adalah sebesar 0,558. Sedangkan Nilai *Silhouette Coefficient* dengan metode *K-Medoids* adalah sebesar 0,529 yang menyatakan bahwa struktur *cluster* yang dihasilkan pada kedua metode pengelompokan ini adalah sama-sama *medium structure* atau terstruktur dengan baik. Dengan demikian hasil penelitian menggunakan jumlah data sebesar 42 data menunjukkan bahwa metode *K-Means* menghasilkan nilai *SC* lebih besar dari pada *K-Medoids* sehingga *K-Means* dapat memberikan hasil pengelompokan yang lebih baik.

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan, maka kesimpulan yang dapat diambil adalah sebagai berikut:

1. Hasil pengelompokan yang terbentuk dengan menggunakan algoritma *K-Means* yaitu untuk *cluster* 1 beranggotakan 37 wilayah, sedangkan untuk *cluster* 2 sebanyak 5 wilayah.
2. Hasil pengelompokan yang terbentuk dengan menggunakan algoritma *K-Medoids* yaitu untuk *cluster* 1 beranggotakan 36 wilayah, sedangkan untuk *cluster* 2 sebanyak 6 wilayah.
3. Nilai *Silhouette Coefficient* dengan metode *K-Means* adalah sebesar 0,558. Sedangkan Nilai *Silhouette Coefficient* dengan metode *K-Medoids* adalah sebesar 0,529 yang menyatakan bahwa metode *K-Means* menghasilkan nilai *SC* lebih besar dari pada *K-Medoids*, sehingga *K-Means* dapat memberikan hasil pengelompokan yang lebih baik.

Daftar Pustaka

Han, J., & M. Kamber. (2006). *Data Mining : Concept and Techniques Second Edition*.

San Francisco, CA: Morgan Kaufmann Publishers.

Kaufman, L., & P. J, Rousseeuw. (1990). *Finding Groups in Data*. New York: John Wiley & Sons.

Kuncoro, M. (2003). *Metode Riset untuk Bisnis dan Ekonomi*. Jakarta: Erlangga.

LAPAN. (2018, Februari 10). [pdf] Tersedia di: <http://modis-catalog.lapan.go.id/monitoring>.

Miller, J. H. & J. Han. (2009). *Geographic Data Mining and Knowledge Discovery*. Boca Raton: Taylor & Francis Group.

NASA. (2018, Maret 18). *VIIRS I-Band 375 m Active Fire Data*. <https://earthdata.nasa.gov/earth-observation-data/near-realtime/firms/active-fire-data>.

Pramesti, D. F., M. T. Furqon, & D. Candra. (2017). Implementasi Metode *K-Medoids* Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 1(9), 723-732.

Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI.

Santoso, S. (2015). *Menguasai Statistik Multivariat Konsep Dasar dan Aplikasi dengan SPSS*. Jakarta: PT Elex Media Komputindo.

Suyanto. (2017). *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Penerbit INFORMATIKA.

Transtoto. (2009). *Kepedulian yang Terganjil*. Jakarta: PT Elex Media Komputindo.

