

# Implementasi *K-Means Clustering* untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model *CRISP-DM*

**F N Dhewayani<sup>1</sup>, D Amelia<sup>2</sup>, D N Alifah<sup>3</sup>, B N Sari<sup>4</sup>, M Jajuli<sup>5</sup>**

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang<sup>12345</sup>

Jl. HS.Ronggo Waluyo, Puseurjaya, Kec. Telukjambe Timur, Kab. Karawang, Jawa Barat<sup>12345</sup>

fitriana.dhewayani18115@student.unsika.ac.id<sup>1\*</sup>, dea.amelia18151@student.unsika.ac.id<sup>2</sup>,  
dinda.nuralifah18044@student.unsika.ac.id<sup>3</sup>, betha.nurina@staff.unsika.ac.id<sup>4</sup>,  
mohammad.jajuli@staff.unsika.ac.id<sup>5</sup>

diterima: 24 Februari 2022

direvisi: 21 Maret 2022

dipublikasi: 31 Maret 2022

## Abstrak

Frekuensi tingginya potensi bahaya bencana yang kerap terjadi di Indonesia yaitu bencana kebakaran hutan dan juga lahan atau biasa disebut dengan Karhutla. Dampak yang terjadi akibat bencana ini sangat besar dan membahayakan, akibat dari bencana kebakaran hutan serta lahan dapat mempengaruhi udara, air, lahan dan juga menyebabkan kerugian finansial, kerusakan fasilitas dan tempat hidup flora dan juga fauna yang ada, serta dapat mengancam nyawa manusia. Informasi daerah bencana dengan potensi kejadian yang tinggi dapat digunakan sebagai informasi agar masyarakat lebih awas dan juga siaga dalam mengenal lingkungan tersebut. Penggunaan teknik clustering dengan menggunakan algoritma K-Means dimana data akan dikelompokkan menjadi beberapa bagian kelompok dan setiap kelompok mempunyai karakteristik yang mirip satu sama lain, tetapi mempunyai karakteristik yang berbeda dengan kelompok lain. Tujuan penelitian dilakukan untuk menghasilkan kelompok daerah yang memiliki potensi tinggi untuk terjadi kebakaran, yang berguna bagi masyarakat agar mengenal bahaya yang mungkin terjadi pada daerah tersebut. Dengan potensi daerah tingkat bencana kebakaran akan ditunjukkan dalam beberapa cluster dan terdapat cluster tertinggi ditujukan kepada Cluster 3 dengan indeks kebakaran di setiap kota tercatat mencapai 87 kejadian setiap tahunnya.

**Kata kunci:** Kebakaran; *Clustering*; *Algoritma K-Means*

## Abstract

*The frequency of potential disaster hazards that often occur in Indonesia is forest and land disasters or known as Karhutla. The impact caused by this disaster is very large and dangerous, the consequences of forest and land fires can affect the air, land and also cause financial losses, damage to facilities and habitats for flora and fauna that exist, and can threaten human life. Information on disaster areas with potential events that can be used as information so that people are more alert and alert in getting to know the environment. The use of clustering techniques using the K-Means algorithm where the data will be divided into several groups and each group has characteristics similar to other groups, but has different characteristics from other groups. The purpose of this research is to produce a group of areas that have a high potential for fires to occur, which are useful for the community to recognize the dangers that may occur in the area. With the potential for fire disaster areas to be shown in several clusters and there is the highest cluster aimed at Cluster 3 with a fire index in each city, 87 incidents are recorded every year.*

**Keywords:** *Fire; Clustering; Algorithm K-means*

## 1. Pendahuluan

Tingginya potensi bahaya bencana yang kerap terjadi di Indonesia yaitu bencana kebakaran adalah bencana kebakaran hutan dan juga lahan atau biasa disebut dengan Karhutla tercatat sudah terjadi sejak tahun 1970 [1]. Dilansir berdasarkan sumber *worldbank* bencana kebakaran hutan dan lahan ini mulai memasuki kejadian skala besar yang terjadi secara tahunan dan diiringi juga dengan rusaknya lingkungan dari dampak ekonomi pada tahun 1982. Dampak yang terjadi akibat bencana kebakaran sangat besar dan membahayakan, akibat dari bencana kebakaran hutan dan lahan dapat mempengaruhi kualitas udara, air, lahan dan juga menyebabkan kerugian finansial, kerusakan fasilitas dan tempat hidup flora dan juga fauna yang ada, kerugian sosial, serta dapat berakibat fatal hingga terjadi kematian [2]. Kebakaran hutan dan lahan juga mengakibatkan kerugian berkurangnya jumlah hutan (deforestasi) dan juga hasil hutan dan berakibat pada pengurangan kualitas potensi desa yang terjadi kebakaran [3]. Hal ini tentu saja dilakukan untuk membuat masyarakat lebih membuka mata dan sadar bahwa bencana kebakaran hutan dan lahan merupakan bencana yang berbahaya dan harus mendapatkan perhatian lebih agar kejadiannya dapat ditanggulangi dengan baik dan ditekan frekuensi kejadiannya karena kerugian yang disebabkan sangat tinggi dan berbahaya.

Menurut Badan Penanggulangan Bencana Daerah (BPBD) (2019), kejadian jumlah bencana kebakaran hutan dan lahan di Jawa Barat merupakan bencana yang tercatat mengalami kejadian paling banyak pada tahun 2019. Kebakaran hutan merupakan bencana terbanyak yang terjadi di Jawa Barat sepanjang tahun 2019, dan kebakaran rumah yang termasuk ke dalam bencana kebakaran lahan juga berada di tempat kedua pada bencana yang terjadi di Jawa Barat. Kebakaran juga terjadi karena beberapa faktor, yaitu faktor alam dan juga manusia [4]. Faktor alam biasanya disebabkan oleh peristiwa alami yang memicu terjadinya kebakaran seperti kebakaran hutan yang dapat diperparah saat musim kemarau datang [5]. Pentingnya kesadaran warga dalam menjaga hutan tentu saja merupakan hal yang sangat penting dalam menanggulangi bencana kebakaran hutan di daerah Jawa Barat yang merupakan daerah dengan jumlah penduduk yang cukup tinggi. Kebakaran lahan yang juga terdapat pada posisi kedua banyaknya kejadian bencana di Jawa Barat juga terjadi akibat dua faktor tersebut yaitu alam dan juga manusia. Penyampaian informasi yang disampaikan oleh Badan Penanggulangan Bencana Daerah (BPBD) masih bersifat acak serta tidak mudah dipahami membuat efektivitasnya belum terlalu dirasakan menjadi masalah karena masyarakat masih memiliki banyak kendala [6].

Pada penelitian sebelumnya penggunaan algoritma *K-Medoids* digunakan untuk pengelompokan wilayah yang memiliki potensi kebakaran tinggi berdasarkan titik persebaran panas [7]. Sedangkan pada penelitian lain dengan teknik clustering yang sama juga digunakan algoritma *K-Means* digunakan untuk melakukan pengelompokan terhadap data penggunaan narkoba dengan metode *elbow* untuk menghasilkan cluster yang paling optimal [8]. Pada penelitian yang dilakukan menghasilkan cluster terbaik dengan menggunakan metode *elbow* tersebut. Selanjutnya pada penelitian terdahulu yang dilakukan untuk melakukan perbandingan hasil clustering antara algoritma *K-Medoids* dan *K-Means* pada pengelompokan kelayakan puskesmas di DIY tahun 2015 didapatkan hasil algoritma *K-Means* lebih baik dari algoritma *K-Medoids* karena memiliki rasio simpangan baku yang lebih kecil [9].

Berdasarkan masalah yang telah disebutkan tersebut pada penelitian ini akan dilakukan pengelompokan daerah rawan bencana kebakaran hutan dan lahan dengan menggunakan *clustering* yang merupakan salah satu metode pada data *mining* yang digunakan untuk pengelompokan data menjadi beberapa *cluster* [10]. Menggunakan algoritma *K-Means* untuk melihat frekuensi terjadinya kebakaran pada setiap daerahnya dengan penggunaan

*Silhouette Coefficient* untuk penentuan *cluster* optimal. Dengan hasil yang akan didapatkan berupa kelompok daerah-daerah yang sebelumnya sudah pernah terjadi bencana kebakaran selama kurun waktu 2019 – 2021 di daerah Jawa Barat, daerah tersebut akan dibagi menjadi beberapa kelompok sesuai dengan frekuensi tingginya kebakaran yang terjadi pada daerah tersebut yang dimana akan berguna bagi masyarakat agar lebih mudah mengetahui daerah mana yang memiliki potensi tinggi terjadi bencana kebakaran. Dimana hasil penelitian ini dapat membuat masyarakat lebih mengenal bahaya yang mungkin terjadi pada lingkungan tersebut yang dapat berakibat fatal dan mengancam nyawa [11].

## 2. Kajian Pustaka

### 2.1 Definisi Kebakaran

Kebakaran merupakan sebuah reaksi oksidasi yang prosesnya cepat dari suatu bahan dengan disertai munculnya api. Berdasarkan sumber yaitu NFPA (*National Fire Protection Association*) kebakaran merupakan terjadinya oksidasi yang melibatkan oksigen, bahan bakar dan sumber panas dan berujung menyebabkan kerugian fisik maupun material [12]. Kebakaran terbagi menjadi beberapa jenis atau diklasifikasikan menjadi beberapa golongan, yang bertujuan untuk mempermudah pemadaman saat bencana terjadi. Klasifikasi kebakaran yaitu sebagai berikut :

1. Kelas A, kebakaran yang terjadi pada barang yang mudah terbakar seperti kertas, plastik dan kayu.
2. Kelas B, kebakaran yang terjadi pada cairan yang mudah terbakar seperti bensin, minyak tanah, dan bahan bakar serupa lainnya.
3. Kelas C, kebakaran yang terjadi pada listrik.
4. Kelas D, kebakaran yang terjadi pada logam yang mudah terbakar seperti titanium, aluminium, kalium dan magnesium.

### 2.2 Faktor terjadi kebakaran

Kebakaran terjadi karena ada pemicu sehingga timbulnya nyala api dan tidak terjadi secara tiba-tiba. Unsur pemicu terbentuknya api yaitu [13]:

1. Bahan yang rentan terbakar, baik dalam bentuk padat (kertas, plastik, kayu, kulit), bahan cair (minyak tanah, bensin, alkohol, cat), atau bahan gas (asetilen, gas alam, butan, propan).
2. Sumber panas yang menjadi pemicu api karena adanya kenaikan suhu sehingga mencapai suhu pembakaran. Contohnya yaitu sinar matahari.
3. Oksigen (%), makin besar kadar oksigen maka nyala api akan semakin besar. Pada kadar oksigen yang kurang dari 12% tidak akan menimbulkan kebakaran,

### 2.3 Data mining

*Data mining* merupakan proses penggalian informasi dari *dataset* menggunakan algoritma dan teknik tertentu yang mempengaruhi disiplin ilmu lain seperti kecerdasan buatan, *machine learning*, statistik, dan *database*. *Data mining* digunakan pada proses pencarian suatu pola tertentu atau informasi pada suatu data menggunakan suatu teknik tertentu. Penggunaan *data mining* biasanya ditujukan untuk melakukan ekstraksi informasi penting pada suatu *dataset* yang besar, dimana *data mining* akan menghasilkan suatu keputusan akhir setelah dilakukan pengolahan data pada kumpulan data yang berjumlah besar. *Data mining* ini berguna untuk membantu dalam pengambilan keputusan di masa yang akan mendatang [14].

### 2.4 Algoritma K-Means

Algoritma K-means merupakan teknik data mining yaitu metode *clustering* atau pengelompokan yang proses pemodelannya tanpa supervisi/pembelajaran serta metode pengelompokan datanya dilakukan secara partisi. Pada metode yang digunakan dalam algoritma K-means, data akan dikelompokkan menjadi beberapa bagian kelompok dan setiap kelompok memiliki ciri-ciri yang mirip dengan satu sama lain, namun memiliki ciri-ciri yang berbeda dengan kelompok lain. Hal itu bertujuan untuk meminimalisir perbedaan antara satu *cluster* dan memaksimalkan perbedaan data dengan *cluster* lain [15]. Berikut ini istilah dalam algoritma k-means:

1. *Cluster*, merupakan suatu kelompok atau grup.
2. *Centroid*, merupakan titik pusat untuk menentukan *Euclidean distance*.
3. Iterasi, merupakan pengulangan suatu proses yang akan berhenti ketika hasil iterasi telah konvergen.

### 2.5 Silhouette coefficient

Metode *Silhouette coefficient* merupakan sebuah metode pengujian kualitas *cluster* yang berfungsi untuk melakukan identifikasi derajat kepemilikan pada setiap objek dalam suatu *cluster*. *Silhouette coefficient* juga merupakan salah satu metode integrasi validasi *clustering* yaitu metode *cohesion* yang berfungsi untuk mengetahui keterkaitan antar objek dalam *cluster* dan *separation* yang berfungsi untuk mengetahui jauhnya jarak yang terpisah antar *cluster* [14].

Dalam menghitung nilai *silhouette coefficient* dilakukan dengan menghitung jarak antar objek dengan *cluster* dengan menggunakan perhitungan *euclidean distance*. Berikut tahapan dari menghitung nilai *silhouette coefficient*:

1. Menghitung nilai rata-rata pada jarak objek data  $i$  dengan seluruh objek data yang berada pada satu *cluster* yang sama.

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

2. Menghitung rata-rata pada objek data  $i$  dengan semua objek data yang ada di klaster lainnya, setelah itu mengambil nilai rata-rata terkecil yang dijadikan sebagai  $b_i$ . Rumus dalam mencari nilai  $b_i$  dapat dilihat pada persamaan (2) berikut.

$$b(i) = \min(D(i, C)) \quad (2)$$

3. Langkah terakhir yaitu menghitung nilai *silhouette coefficient*

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

Hasil dari perhitungan nilai *silhouette coefficient* memiliki nilai yang bervariasi dengan rentang -1 sampai 1. Nilai suatu *clustering* dikatakan baik jika memiliki nilai *silhouette coefficient* bernilai positif ( $a(i) < b(i)$ ) dan  $a(i)$  mendekati nilai 0. Dengan begitu menghasilkan nilai *silhouette coefficient* yang maksimal bernilai 1 ketika  $a(i) = 0$ . Jika nilai *silhouette coefficient* bernilai 1 menandakan bahwa *cluster* 1 berada pada *cluster* yang tepat. Tetapi, jika nilai *silhouette coefficient* bernilai 0 maka objek  $i$  berada pada *cluster* yang memiliki struktur yang tidak jelas. Namun, jika nilai *silhouette coefficient* memiliki nilai -1, maka struktur dari *cluster* memiliki nilai *overlapping*, sehingga objek  $i$  lebih tepat dimasukkan pada klaster lain. Berikut ini nilai *silhouette coefficient* menurut Rousseeu (1987) [16] :

- a.  $0.7 < SC \leq 1$ , Struktur *cluster* yang kuat (*Strong structure*)

- b.  $0.5 < SC \leq 0.7$ , Struktur *cluster* yang standard (*Medium structure*)
- c.  $0.25 < SC \leq 0.5$ , Struktur *cluster* yang lemah (*Weak structure*)
- d.  $SC \leq 0.25$ , Tidak memiliki struktur (*No structure*)

## 2.6 CRISP-DM

CRISP-DM (*Cross Industry Standard for Data mining*) merupakan metode yang banyak digunakan oleh para ahli dengan menggunakan proses pemodelan data di dalamnya, Tujuan dari metode CRISP-DM ini yaitu untuk menemukan pola yang menarik dan memiliki makna pada data yang digunakan. CRISP-DM memiliki tahapan dan kerangka kerja yang terstruktur sehingga pengguna metode ini akan lebih terarah dan mengetahui langkah yang harus dikerjakan dalam penelitian. Pada CRISP-DM ini terdapat 6 tahap [17], yaitu :

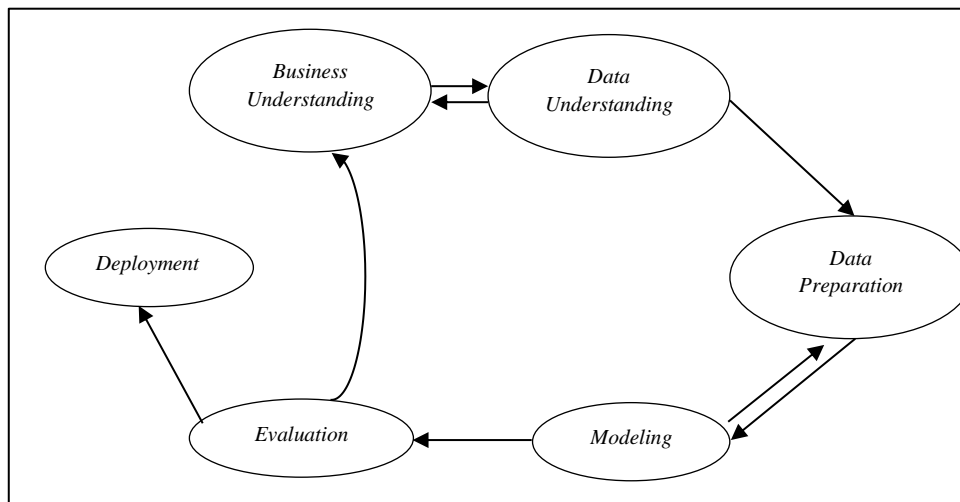
1. *Business Understanding*  
*Business Understanding* adalah proses menentukan tujuan bisnis, memahami situasi dan kondisi pada saat penelitian dan menetapkan sebuah tujuan dari penelitian yang dilakukan ke dalam permasalahan yang diselesaikan dengan *data mining*.
2. *Data Understanding*  
*Data understanding* adalah tahap persiapan, melakukan pengecekan terhadap data yang digunakan, mengumpulkan data awal serta melakukan identifikasi pada kualitas data. Dalam *data understanding*, data yang digunakan akan melalui proses deskripsi dari setiap fiturnya.
3. *Data Preparation*  
*Data preparation* merupakan proses yang dilakukan setelah data telah dikumpulkan. Pada tahap ini, data akan melalui proses identifikasi, pemilihan data, pembersihan data dan transformasi data.
4. *Modeling*  
*Modeling* merupakan tahap implementasi algoritma yang akan digunakan untuk melakukan pencarian, identifikasi, serta menghasilkan pola yang akan digunakan pada data penelitian.
5. *Evaluation*  
*Evaluation* adalah suatu proses untuk melakukan pengukuran hasil evaluasi dari model yang telah diimplementasikan sebelumnya di tahap *modelling*. Hasil evaluasi tersebut menggambarkan proses dari *data mining* yang telah dilakukan dan mengukur model yang paling baik untuk digunakan.
6. *Deployment*  
*Deployment* adalah proses pembuatan laporan atau artikel jurnal menggunakan hasil penelitian.

## 2.7 Python

*Python* adalah salah satu bahasa pemrograman tingkat tinggi yang dapat secara langsung menjalankan sekaligus instruksi multi-guna menggunakan pendekatan berorientasi objek, serta menggunakan sistem *semantic* dinamis untuk memberikan tingkat tertentu dari keterbacaan sebuah *syntax*. *Python* juga mudah dipelajari karena telah dilengkapi oleh manajemen memori otomatis (*pointer*) [18].

## 3. Metode Penelitian

Penelitian ini menggunakan salah satu teknik data mining yaitu clustering dengan algoritma K-means. Metode CRISP-DM memiliki enam tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.



Gambar 1. Metode CRISP-DM

### 3.1 Business Understanding

Pada tahap pertama yaitu dilakukan pemahaman bisnis serta penjelasan terhadap masalah *data mining* sehingga dapat mencapai tujuan. Pada tahap *business understanding* dilakukan proses penentuan tujuan bisnis, memahami situasi dan kondisi pada saat penelitian dan menetapkan sebuah tujuan dari penelitian yang dilakukan ke dalam permasalahan yang diselesaikan dengan *data mining*.

### 3.2 Data Understanding

Tahap *data understanding* merupakan tahap persiapan, dilakukan pengecekan terhadap data yang digunakan, mengumpulkan data awal serta melakukan identifikasi pada kualitas data. Dalam *data understanding*, data yang digunakan akan melalui proses deskripsi dari setiap fiturnya.

### 3.3 Data Preparation

Pada *data preparation* dilakukan persiapan data dengan menyesuaikan *dataset* agar dapat sesuai dengan kebutuhan yang akan digunakan saat tahap pemodelan. Pada tahap *data preparation* pertama akan dilakukan *data selection* untuk melakukan seleksi terhadap data yang dibutuhkan, *data preprocessing* untuk mempersiapkan data mentah sehingga siap untuk menjadi data yang dapat digunakan pada pemodelan, dan *data transformation* untuk melakukan transformasi data.

### 3.4 Modeling

Pada tahap ini, pemodelan menggunakan teknik *clustering* untuk mendapatkan suatu pola *data mining*. Teknik pemodelan *clustering* yang digunakan adalah algoritma k-means. Sebelum melakukan pemodelan, langkah awal yaitu menentukan *cluster*. Untuk menentukan jumlah *cluster* terbaik, peneliti menggunakan metode *silhouette coefficient*.

### 3.5 Evaluation

Setelah melakukan tahap pemodelan, tahap selanjutnya adalah evaluasi untuk mengukur model yang paling baik untuk digunakan. Evaluasi yang digunakan yaitu dengan teknik *silhouette score*.

### 3.6 Deployment

Setelah tahap evaluasi selesai dilakukan untuk menilai secara detail hasil dari sebuah model maka dilakukan implementasi dari keseluruhan model yang telah dibangun. Pada penelitian kali ini implementasi yang dilakukan akan diterapkan dan dibuat menjadi sebuah laporan yang berisi hasil yang sesuai dengan tahap CRISP-DM.

#### 4. Hasil dan Pembahasan

##### 4.1 Business Understanding

Pada tahap pertama yaitu dilakukan pemahaman bisnis serta penjelasan terhadap masalah *data mining* sehingga dapat mencapai tujuan. Pemahaman masalah pada penelitian ini mengacu pada kebakaran di daerah Jawa Barat. Kebakaran menjadi salah satu bencana yang tidak terduga dan dapat datang kapan dan dimana saja, baik kebakaran lahan maupun kebakaran bangunan. Jawa Barat menjadi salah satu provinsi yang banyak penduduk dan lahannya, sehingga resiko kebakaran kemungkinan besar dapat terjadi. Untuk itu, untuk mengetahui titik rawan tingginya bencana kebakaran, perlu dilakukannya pengelompokan daerah rawan bencana kebakaran agar pemerintah Jawa Barat melakukan tindakan preventif sebagai langkah awal meminimalisir bencana kebakaran.

##### 4.2 Data Understanding

*Dataset* yang digunakan pada penelitian ini yaitu data kebakaran tahun 2019-2021 Provinsi Jawa Barat. Sumber *dataset* yaitu berasal dari [opendata.jabarprov.go.id](http://opendata.jabarprov.go.id). *Dataset* kebakaran lahan dan kebakaran hutan terdiri dari 135 records yaitu berisi nama-nama kabupaten/kota dan disertai 9 atribut, yaitu id, kode provinsi, nama provinsi, nama kabupaten kota, kode kabupaten kota, jumlah kebakaran bangunan, jumlah kebakaran lahan, satuan dan tahun. Berikut ini *dataset* yang digunakan pada penelitian yang ditunjukkan pada Tabel 1.

Tabel 1. Dataset Awal

Id	Kode Provinsi	Kode Kab/Kota	Nama Kab/Kota	Jumlah Kebakaran Bangunan	Jumlah Kebakaran Lahan	Satuan	Tahun
1	32	3201	Bogor	1	5	Kejadian	2019
2	32	3202	Sukabumi	61	32	Kejadian	2019
3	32	3203	Cianjur	3	7	Kejadian	2019
4	32	3204	Bandung	25	22	Kejadian	2019
5	32	3205	Garut	17	36	Kejadian	2019
..	..	..	..	..	..	..	..
135	32	3279	Banjar	0	0	Kejadian	2021

##### 4.3 Data Preparation

Pada *data preparation* adalah proses melakukan persiapan data dengan menyesuaikan *dataset* agar dapat sesuai dengan kebutuhan yang akan digunakan saat tahap pemodelan.

###### 1. Data Selection

*Dataset* kebakaran yang didapatkan pada *dataset* 2019-2021 dengan banyak 135 data. Kemudian data ini akan diseleksi, dari total terdapat 9 atribut hanya 6 atribut saja yang dipilih yaitu id, kode provinsi, kode kabupaten kota, jumlah kebakaran bangunan, jumlah kebakaran lahan, dan tahun. Data hasil seleksi dapat dilihat pada Tabel 2 berikut.

Tabel 1. Hasil *Data Selection*

Id	Kode Provinsi	Kode Kab/Kota	Nama Kab/Kota	Jumlah Kebakaran Bangunan	Jumlah Kebakaran Lahan	Tahun
1	32	3201	Bogor	1	5	2019
2	32	3202	Sukabumi	61	32	2019
3	32	3203	Cianjur	3	7	2019
4	32	3204	Bandung	25	22	2019
5	32	3205	Garut	17	36	2019
..	..	..	..	..	..	..
135	32	3279	Banjar	0	0	2021

Pada Tabel 2. *Dataset* awal yang tidak dibutuhkan akan dihapus karena tidak sesuai dengan kebutuhan, yaitu nama provinsi, nama kabupaten kota dan satuan.

## 2. *Data Preprocessing*

Pada tahap *data preprocessing* akan dilakukan persiapan data mentah sehingga siap untuk menjadi data yang dapat digunakan pada pemodelan. Tahap *preprocessing* biasanya berupa proses pembersihan data, salah satunya menangani *missing value*. Pada *dataset* yang digunakan pada penelitian ini lengkap dan tidak terdapat *missing value* maupun data yang inkonsisten, sehingga dapat lanjut ke tahap berikutnya yaitu tahap *data transformation*.

## 3. *Data Transformation*

Tahap selanjutnya yaitu tahap transformasi data. Transformasi data ini adalah mengubah standar angka pada data menjadi nilai 0-1. Teknik transformasi yang digunakan adalah *Standard Scaler*. *Standar Scaler* bertujuan membuat rata-rata 0 dan variansi 1.

```

X = df[['jumlah_kebakaran_bangunan', 'jumlah_kebakaran_lahan']]
sc = StandardScaler()
sc.fit(X)
X = sc.transform(X)
x|

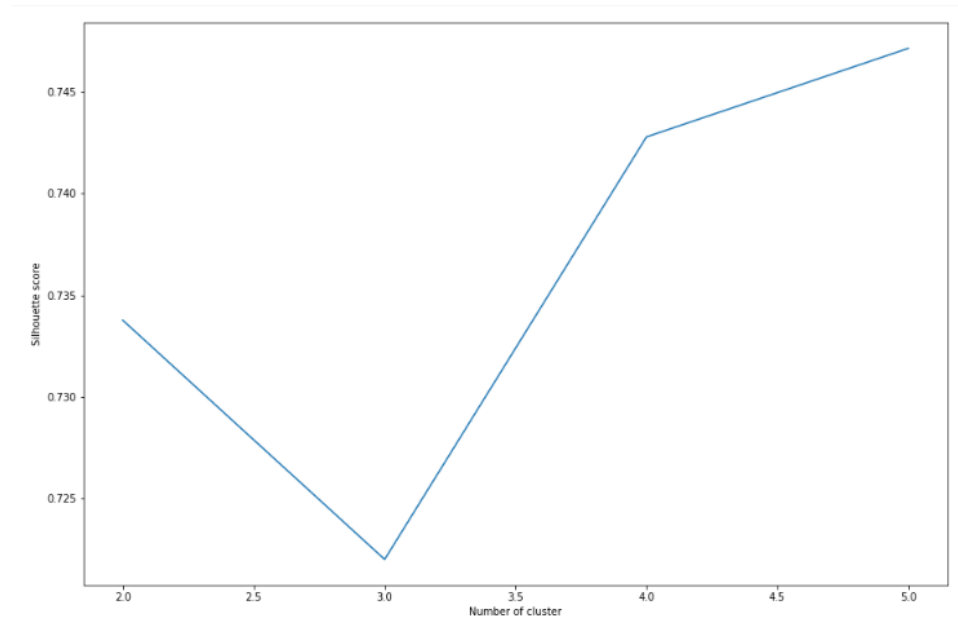
```

Gambar 2. Proses Transformasi Data

## 4.4 *Modeling*

Pada tahap ini, pemodelan menggunakan teknik *clustering* untuk mendapatkan suatu pola *data mining*. Teknik pemodelan *clustering* yang digunakan adalah algoritma k-means. Sebelum melakukan pemodelan, langkah awal yaitu menentukan *cluster*. Untuk menentukan jumlah *cluster* terbaik, peneliti menggunakan metode *silhouette coefficient* seperti pada Gambar 3 berikut.





Gambar 3. Grafik *Silhouette coefficient*

Dapat dilihat hasil grafik pada Gambar 3, setelah melakukan teknik *silhouette coefficient* didapatkan bahwa *cluster=5* memiliki nilai rata-rata nilai *silhouette coefficient* paling tinggi diantara yang lainnya yaitu sebesar 0.7483035806217097. Sehingga, pada penelitian ini menggunakan *cluster=5* sebagai nilai *k* untuk pemodelan *k-means*.

Setelah menentukan *cluster*, tahap berikutnya adalah pemodelan dengan menggunakan algoritma *K-means*. Berikut proses *k-means clustering* pada *dataset* kebakaran tahun 2019-2021 dapat dilihat pada Gambar 4 berikut.

```

km = KMeans(n_clusters = 5, init='k-means++', random_state=0)
km.fit(X)
centroids = km.cluster_centers_
X['Labels'] = km.labels_

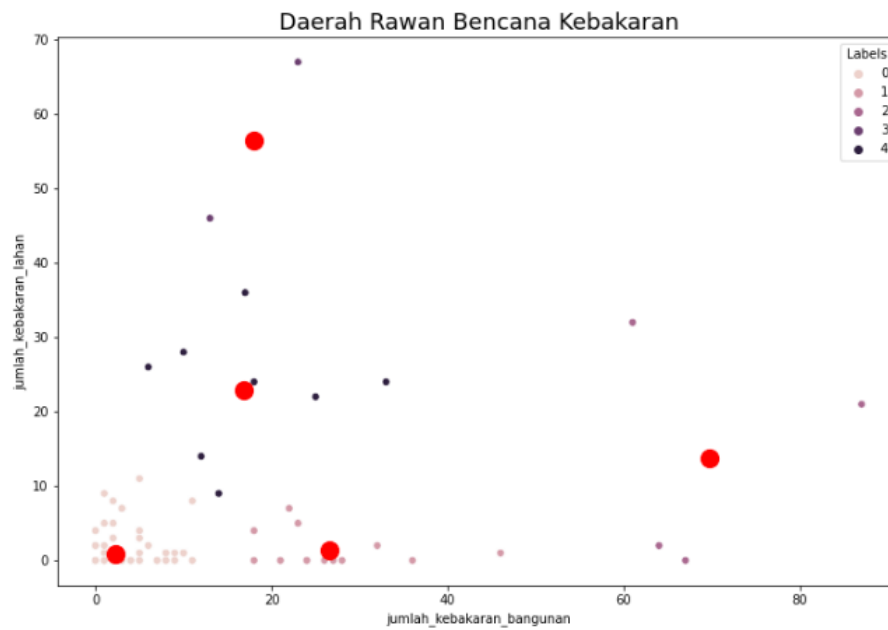
plt.figure(figsize=(12,8))
sns.scatterplot(X['jumlah_kebakaran_bangunan'], X['jumlah_kebakaran_lahan'], hue=X['Labels'])

plt.scatter(centroids[:,0], centroids[:,1], c='red', s=200)
plt.title('Daerah Rawan Bencana Kebakaran', fontsize=18)
plt.show()

```

Gambar 4. *Source Code* Pemodelan *K-Means*

Kemudian setelah tahap pemodelan *cluster* dilakukan dengan menggunakan algoritma *k-means* yang ada pada Gambar 4 dihasilkan kelompok *cluster* pada Gambar 5 grafik pengelompokkan *cluster*. Terdapat 5 titik kelompok *cluster* yang terbentuk dari hasil pemodelan tersebut dengan setiap titik *cluster* yang terbentuk ditandai dengan titik merah bulat besar dan setiap anggota *cluster* di tandai dengan titik kecil berbagai warna di sekitarnya yang merupakan penyebaran daerah yang terdapat di setiap kelompok.



Gambar 5. Grafik Pengelompokan *Cluster*

Setelah melakukan pemodelan *clustering* dengan k-means, hasil *cluster* dari setiap data telah diketahui. Didapatkan data jumlah tiap pengelompokan ditampilkan pada Tabel 3 berikut.

Tabel 3. Pengelompokan K-Means

Jumlah <i>Cluster</i>	Jumlah Anggota <i>Cluster</i>
2	<i>Cluster</i> 1(21 data) <i>Cluster</i> 2(114 data)
3	<i>Cluster</i> 1(108 data) <i>Cluster</i> 2(5 data) <i>Cluster</i> 3(22 data)
4	<i>Cluster</i> 1(13 data) <i>Cluster</i> 2(110 data) <i>Cluster</i> 3(8 data) <i>Cluster</i> 4(4 data)
5	<i>Cluster</i> 1(108 data) <i>Cluster</i> 2(13 data) <i>Cluster</i> 3(4 data) <i>Cluster</i> 4(2 data) <i>Cluster</i> 5(8 data)
6	<i>Cluster</i> 1(90 data) <i>Cluster</i> 2(7 data) <i>Cluster</i> 3(4 data) <i>Cluster</i> 4(21 data) <i>Cluster</i> 5(11 data) <i>Cluster</i> 6(2 data)

Pengujian jumlah *Cluster* berguna untuk mengetahui jumlah *Cluster* terbaik dengan cara melihat hasil dari nilai *Silhouette coefficient* yang dihasilkan dari pengujian yang telah

dilakukan pada 135 data jumlah bencana kebakaran. Dengan jumlah nilai *Cluster* (*k*) yang diujikan adalah *k* 2 sampai dengan 6. Hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Hasil Pengujian Jumlah *Cluster* (*k*)

Data Uji	Nilai <i>k</i>	Nilai Rata - Rata <i>Silhouette coefficient</i>
131	2	0.73339624117858
	3	0.7229095488437975
	4	0.7433271508345515
	<b>5</b>	<b>0.7483035806217097</b>
	6	0.7387868232116501

Hasil pengujian jumlah *Cluster* menunjukkan masing-masing nilai yang diperoleh setiap *Cluster*, hasil nilai tertinggi didapatkan pada penggunaan nilai *Cluster* *k*=5 dengan nilai rata-rata *Silhouette coefficient* 0.7483035806217097.

Setelah dilakukan pengujian secara keseluruhan, diperoleh hasil nilai rata-rata *Silhouette coefficient* yang terdapat pada Tabel 5, menunjukkan nilai tertinggi didapatkan oleh *k*=5 yaitu 0.7483035806217097. Hasil pengujian nilai rata-rata *Silhouette coefficient* yang didapatkan oleh *k*=5 menunjukkan bahwa nilai *k* tersebut merupakan *Cluster* optimal yang dapat menghasilkan informasi terbaik. *Cluster* 5 merupakan *Cluster* yang menampilkan hasil terbaik dari pengujian yang dilakukan dari uji *k*=2 sampai *k*=6, memiliki jumlah anggota masing-masing *Cluster* adalah 108, 13, 4, 2 dan 8 data. Berikut merupakan hasil informasi yang diperoleh dari *Cluster* 5 pada Tabel 5.

Tabel 5. Informasi Hasil *Cluster*

<i>Cluster</i>	Daerah Rawan	Karakteristik Kejadian	Kriteria
<i>Cluster</i> 1 (108 data)	Bogor, Cianjur, Tasikmalaya, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Sukabumi, Cirebon, Depok, Banjar, Bandung, Garut, Kuningan, Majalengka, Pangandaran, Cimahi, Ciamis, Sumedang	0-11	Sangat Rendah
<i>Cluster</i> 2 (13 data)	Ciamis, Cirebon, Garut, Majalengka, Pangandaran, Sukabumi, Sumedang, Bogor, Cimahi	0-46	Sedang
<i>Cluster</i> 3 (4 data)	Sukabumi, Sumedang, Bandung	0-87	Sangat Tinggi
<i>Cluster</i> 4 (2 data)	Majalengka, Sumedang	0-67	Tinggi
<i>Cluster</i> 5 (8 data)	Bandung, Ciamis, Cirebon, Garut, Kuningan, Pangandaran, Cimahi	0-33	Rendah

Pada Tabel 5. terdapat hasil informasi cluster yang telah menunjukkan daerah mana saja yang termasuk kedalam cluster penyebaran jumlah daerah sesuai dengan cluster masing-masing berdasarkan hasil pengujian. Terdapat 5 kriteria yang digunakan untuk melihat hasil perhitungan, yaitu : Sangat rendah terdapat 0-11 kejadian kebakaran, rendah terdapat 0-33 kejadian kebakaran, sedang terdapat 0-46 kejadian kebakaran, tinggi terdapat 0-67, dan sangat tinggi 0-87. Kriteria tersebut dapat menunjukkan tingkat frekuensi kebakaran yang

terjadi pada setiap daerah cluster. *Cluster* 3 mendapatkan hasil tertinggi dengan kriteria sangat tinggi sedangkan *cluster* 1 mendapatkan hasil terendah dengan kriteria sangat rendah.

#### 4.5 Evaluation

Setelah melakukan tahap pemodelan, tahap selanjutnya adalah evaluasi. Evaluasi yang digunakan yaitu dengan teknik *silhouette score* dengan proses pada Gambar 6.

```
for n_clusters in range(2, 7):  
  
    clusterer = KMeans(n_clusters=n_clusters,max_iter=1000, random_state=10)  
    cluster_labels = clusterer.fit_predict(X)  
  
    silhouette_avg = silhouette_score(X, cluster_labels)  
    print("Untuk n_clusters =", n_clusters,  
          "Nilai Rata - Rata Silhouette :", silhouette_avg)  
  
Untuk n_clusters = 2 Nilai Rata - Rata Silhouette : 0.73339624117858  
Untuk n_clusters = 3 Nilai Rata - Rata Silhouette : 0.7229095488437975  
Untuk n_clusters = 4 Nilai Rata - Rata Silhouette : 0.7433271508345515  
Untuk n_clusters = 5 Nilai Rata - Rata Silhouette : 0.7483035806217097  
Untuk n_clusters = 6 Nilai Rata - Rata Silhouette : 0.7387868232116501
```

Gambar 6. Proses Evaluasi *Silhouette coefficient*

Pada Gambar 6 dapat dilihat bahwa nilai *silhouette coefficient* pada *cluster*=5 memiliki nilai paling tinggi dibandingkan dengan nilai *cluster* lain dengan rata-rata *silhouette coefficient* yaitu 0.7483035806217097. Kualitas dari *cluster* yang dihasilkan dari algoritma k-means termasuk dalam *strong structure* dengan interpretasi klaster terbaik sudah ditemukan.

#### 4.6 Deployment

Setelah tahap evaluasi selesai dilakukan untuk menilai secara detail hasil dari sebuah model maka dilakukan implementasi dari keseluruhan model yang telah dibangun. Pada penelitian kali ini implementasi yang dilakukan akan diterapkan dan dibuat menjadi sebuah laporan yang berisi hasil yang sesuai dengan tahap CRISP-DM.

### 5. Kesimpulan

Jumlah *cluster* yang dapat memberikan nilai tertinggi terhadap nilai *silhouette coefficient* yaitu *cluster* 5. Hasil tertinggi didapatkan oleh nilai k=5 yaitu 0.7483035806217097 dengan jumlah data uji 135 data jumlah kebakaran bangunan dan lahan. Analisis penentuan potensi daerah dengan tingkat bencana kebakaran menghasilkan 5 *cluster* uji yaitu *cluster* 1 yaitu 108 data, *cluster* 2 dengan 13 data, *cluster* 5 dengan 8 data, *cluster* 3 dengan 4 data, dan *cluster* 5 dengan 8 data. Potensi daerah tingkat bencana kebakaran tertinggi ditujukan kepada *cluster* 3 dengan kriteria sangat tinggi dengan indeks kebakaran di setiap kota tercatat mencapai 87 kejadian setiap tahunnya. Berdasarkan penelitian yang telah dilakukan dapat diambil kesimpulan algoritma *K-Means* telah berhasil melakukan pengelompokan dengan mendapatkan anggota *cluster* yang menghasilkan hasil yang optimal untuk memberikan informasi potensi kebakaran hutan dan lahan yang terdapat di setiap daerah. Kelebihan pada penelitian ini penggunaan nilai rata-rata *silhouette coefficient* yang digunakan untuk penentuan *cluster* terbaik memberikan kemudahan dalam proses penelitian, terdapat jumlah serta nama daerah pada setiap *cluster* serta hasil dapat terlihat pada kriteria pada informasi setiap *cluster*. Namun kekurangan pada penelitian ini terdapat beberapa anggota *cluster* yang tidak berada ditempat yang tepat karena pengujian awal yang tidak dapat ditentukan melaikan langsung berdasarkan python sendiri.

### Daftar Pustaka

- [1] M. Badri, D. P. Lubis, D. Susanto, and D. Suharjito, "Sistem Komunikasi Peringatan Dini Pencegahan Kebakaran Hutan Dan Lahan Di Provinsi Riau," *J. PIKOM (Penelitian Komun. dan Pembangunan)*, vol. 19, no. 1, p. 1, 2018, doi: 10.31346/jpikom.v19i1.1266.
- [2] H. Panja, "Penerapan Sarana Alat Pemadam Api Ringan di Pusat Perbelanjaan Mall," *Higeia J. Public Heal. Res. Dev.*, vol. 4, no. 2, pp. 280–290, 2020.
- [3] A. Primajaya, B. N. Sari, and A. Khusaeri, "Prediksi Potensi Kebakaran Hutan dengan Algoritma Klasifikasi C4.5 Studi Kasus Provinsi Kalimantan Barat," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 2, p. 188, 2020, doi: 10.26418/jp.v6i2.37834.
- [4] A. Syamsuadi, D. Arisandi, T. F. Gafar, L. Trisnawati, S. Hartati, and M. F. Anugerah, "Peningkatan Kewaspadaan Bencana Asap Di Kecamatan Dayun Kabupaten Siak," *Community Dev. J. J. Pengabd. Masy.*, vol. 1, no. 2, pp. 48–51, 2020, doi: 10.31004/cdj.v1i2.701.
- [5] 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.
- [6] T. I. Hermanto, "Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana," *J-SAKTI (Jurnal Sains Komput. dan Inform.)*, vol. 5, pp. 406–416, 2021, [Online]. Available: <http://ejournal.tunasbangsa.ac.id/index.php/jsakti/article/view/332>.
- [7] 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.
- [8] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [9] Y. H. Susanti and E. Widodo, "Perbandingan K-Means dan K-Medoids Clustering terhadap Kelayakan Puskesmas di DIY Tahun 2015," *Pros. SI MaNis (Seminar Nas. Integr. Mat. dan Nilai Islam.)*, vol. 1, no. 1, pp. 116–122, 2017.
- [10] Ahmad Harmain, P. Paiman, H. Kurniawan, K. Kusriani, 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.
- [11] D. Taruna, A. Fauzi, and M. C. Aruan, "Aplikasi Pengenalan Dan Pencegahan Bencana Kebakaran Api Yang Disebabkan Oleh Manusia ( Human Error ) Berbasis Android," *Semna Ristek (Seminar Nas. Ris. dan Inov. Teknol.)*, pp. 1–7, 2021.
- [12] S. R. Ary Firman Sanjoto, Ana Komari, "Analisa Upaya Pencegahan Dan Penanggulangan Potensi Bahaya," vol. 1, no. 1, pp. 24–33, 2019.
- [13] I. Pamungkas, I. Wahyuni, and S. Suroto, "Analisis Kesiapsiagaan Polisi Kehutanan Terhadap Potensi Dan Pencegahan Kebakaran Hutan Di Wilayah Kph Semarang Perum Perhutani Divisi Regional Jawa Tengah (Studi Kasus Di Bagian Kesatuan Pemangkuan Hutan Penggaron)," *J. Kesehat. Masy. Univ. Diponegoro*, vol. 4, no. 4, pp. 690–697, 2016.
- [14] G. N. W. Paramartha, D. E. Ratnawati, and A. W. Widodo, "Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks

- Pembangunan Manusia (IPM),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. Vol. 1, no. 9, pp. 813–824, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>.
- [15] R. Muliono and Z. Sembiring, “Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen,” *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 2502–714, 2019.
- [16] D. S. Utami and D. R. S. Saputro, “Pengelompokan Data Yang Memuat Pencilandengan Kriteria Elbow Dan Koefisien Silhouette (Algoritme K-Medoids),” *Knpmp Iii 2018*, pp. 448–456, 2018.
- [17] 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,” vol. 5, no. 2, 2021.
- [18] I. A. Permatasari, *Klasifikasi Penduduk Penerima Bantuan Covid-19 Di Kecamatan Kasomalang*. 2021.