

IMPLEMENTASI METODE K-MEANS DAN K-MEDOIDS PADA CLUSTERING TINGKAT KEPEDULIAN MASYARAKAT TERHADAP BENCANA KEBAKARAN

Yogi Ersan Fadrial¹, Yogi Yunefri², Sutejo³, Fajrizal⁴, Syahtriatna⁵

(^{1,2,3,4}Program Studi Teknik Informatika, ⁵Program Studi Teknik Informatika Fakultas Ilmu Komputer
Universitas Lancang Kuning)

(Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015)

e-mail: ¹yogiersan@unilak.ac.id, ²yogyunefri@unilak.ac.id, ³sutejo@unilak.ac.id,
⁴fajrizal@unilak.ac.id, ⁵syahtriatna@unila.ac.id

Abstrak

Penelitian ini bertujuan untuk mengelompokkan tingkat kepedulian masyarakat Kota Pekanbaru terhadap bencana kebakaran menggunakan metode klasterisasi K-Means dan K-Medoids. Latar belakang penelitian ini adalah tingginya frekuensi kebakaran di Kota Pekanbaru yang mencapai lebih dari 150 kasus per tahun, menekankan pentingnya kesadaran dan kesiapsiagaan masyarakat dalam mengurangi dampak bencana tersebut. Metode penelitian mencakup pengumpulan data melalui kuesioner berbasis skala Likert, diikuti dengan pra-proses data dan analisis menggunakan metode Elbow untuk menentukan jumlah kluster optimal. Data kemudian dikelompokkan menggunakan algoritma K-Means dan K-Medoids yang diterapkan melalui Google Colaboratory. Hasil analisis menunjukkan bahwa metode K-Means menghasilkan tiga kluster dengan tingkat kepedulian keseluruhan tertinggi pada kluster usia muda (22,76 tahun), sementara metode K-Medoids menunjukkan stabilitas data yang lebih baik dengan nilai rata-rata kepedulian tertinggi sebesar 2,88. Kesimpulannya, klasterisasi menggunakan kedua metode ini memberikan wawasan mengenai pola kesadaran masyarakat, yang dapat menjadi dasar bagi pemerintah dan lembaga terkait dalam merumuskan kebijakan serta program edukasi yang lebih efektif untuk mitigasi kebakaran di masa depan.

Kata kunci: K-Mean K-Medoids, Clustering, kepedulian masyarakat, bencana kebakaran, data mining

Abstract

This study aims to cluster the level of public awareness in Pekanbaru City regarding fire disasters using the K-Means and K-Medoids clustering methods. The background of this study is the high frequency of fires in Pekanbaru City, which reaches more than 150 cases per year, emphasizing the importance of public awareness and preparedness in reducing the impact of such disasters. The research method involves data collection through Likert scale-based questionnaires, followed by data preprocessing and analysis using the Elbow method to determine the optimal number of clusters. The data is then grouped using the K-Means and K-Medoids algorithms implemented through Google Colaboratory. The analysis results show that the K-Means method generates three clusters, with the highest overall awareness level in the younger age group (22.76 years). Meanwhile, the K-Medoids method shows better data stability, with the highest average awareness value of 2.88. In conclusion, clustering using both methods provides insights into the public's awareness patterns, which can serve as a basis for the government and related institutions to formulate policies and educational programs more effectively for fire disaster mitigation in the future.

Keywords: K-Means K-Medoids, Clustering, Community Awareness, fire Disaster, Data Mining

1. PENDAHULUAN

Bencana kebakaran adalah suatu reaksi oksidasi eksotermis yang berlangsung dengan cepat dari suatu bahan bakar yang disertai timbulnya api. Di Indonesia sudah sering terjadi kebakaran pemukiman di kota-kota besar yang padat penduduk (Hamzah et al., 2019). Di berbagai wilayah, termasuk Kota

Pekanbaru, risiko kebakaran menjadi semakin memprihatinkan seiring dengan perkembangan kota dan perubahan lingkungan yang mendukung penyebaran api. Frekuensi kebakaran lahan dan bangunan di Pekanbaru mencapai lebih dari 150 kasus per tahun dan mengalami puncaknya di setiap tahunnya yaitu mencapai kurang lebih 200 kasus (sumber : pekanbaru.go.id). Untuk mengurangi dampak bencana kebakaran dan memitigasi risikonya, penting untuk memahami tingkat kesadaran dan kepedulian masyarakat terhadap bahaya kebakaran.

Besarnya dampak semakin parah di tempat-tempat dengan tingkat kesiapsiagaan bencana kebakaran yang rendah. Kesiapsiagaan bencana salah satu elemen penting dalam pengurangan risiko bencana dan meliputi kesadaran masyarakat, kesiapan memberikan tanggapan yang tepat dan pemulihan yang cepat. Masyarakat memiliki peran dalam mengurangi resiko kebencanaan kebakaran (Trifianingsih et al., 2022). Masyarakat yang sadar dan peduli terhadap risiko kebakaran lebih cenderung mengambil tindakan pencegahan, seperti pemeliharaan peralatan pemadam api dan perencanaan evakuasi yang baik. Selain itu, mereka juga lebih siap untuk merespons dengan cepat saat kebakaran terjadi, yang dapat membantu mengurangi kerugian dan dampak negatif.

Salah satu pendekatan yang dapat digunakan untuk analisis tingkat kepedulian masyarakat adalah klusterisasi data. Klusterisasi merupakan proses pengelompokan beberapa titik dalam dua atau lebih kelompok hingga data dalam kelompok yang sama menjadi mirip satu sama lain berdasarkan informasi yang tersedia dengan poin data (Luchia et al., 2022). Dalam konteks ini, penelitian yang dilakukan akan menggunakan dua metode klusterisasi yaitu K-Means dan K-Medoids untuk mengelompokkan masyarakat Kota Pekanbaru ke dalam kelompok-kelompok berdasarkan tingkat kepedulian mereka terhadap bencana kebakaran.

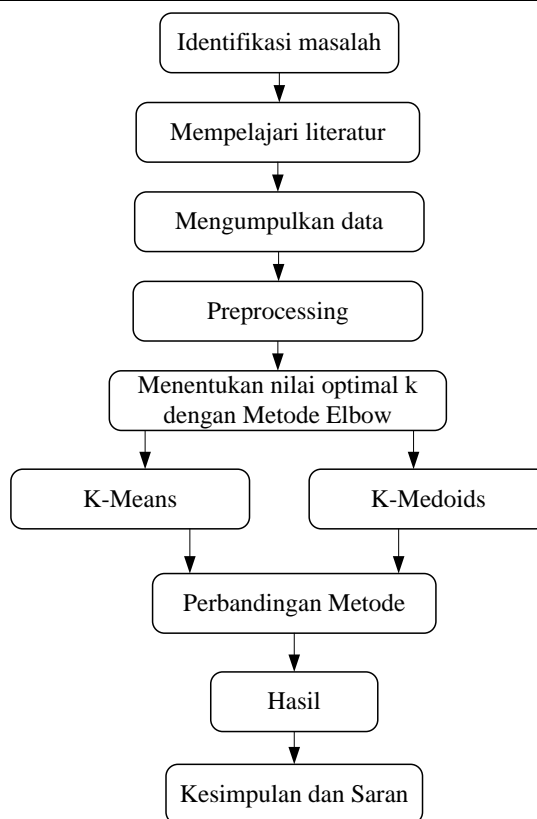
Dengan demikian, penelitian ini bertujuan untuk mengelompokkan pola tingkat kepedulian masyarakat terhadap bencana kebakaran di Kota Pekanbaru. Hasil dari penelitian ini diharapkan dapat memberikan wawasan berharga kepada pemangku kepentingan, termasuk pemerintah daerah dan lembaga kemanusiaan, untuk merumuskan kebijakan yang lebih efektif dalam meningkatkan kesadaran dan kepedulian masyarakat terhadap risiko kebakaran. Sebagaimana yang disebutkan oleh (Marfuah et al., 2020) tentang kegiatan pelatihan pencegahan dan penanganan kebakaran dalam pengabdian masyarakat dapat memberikan pengetahuan mengenai faktor penyebab terjadinya kebakaran dan mengetahui tindakan pencegahan untuk meminimalisir terjadinya kebakaran dalam kegiatan sehari-hari serta dapat mendorong kepedulian terhadap pencegahan dan penanganan kebakaran dalam komunitas lingkungan masyarakat.

Selain itu, penelitian ini juga dapat memberikan dasar untuk pengembangan program-program edukasi dan advokasi yang lebih terfokus dalam meningkatkan respons masyarakat terhadap ancaman bencana kebakaran di masa depan. Berdasarkan dari latar belakang masalah diatas, penulis bermaksud untuk mengangkat penelitian yang berjudul Klusterisasi Tingkat Kepedulian Masyarakat Kota Pekanbaru Terhadap Bencana Kebakaran Dengan Metode K-Means dan K-Medoids.

Adapun tujuan dari penelitian ini adalah untuk melakukan analisis pengelompokkan dengan penerapan *data mining clustering* dalam melihat tingkat kepedulian masyarakat kota Pekanbaru terhadap bencana kebakaran dengan metode K-Means dan K-Medoids yang diuji dengan menggunakan *tools* Google Colaboratory.

2. METODE PENELITIAN

Dalam metodologi penelitian ada tahapan-tahapan penelitian yang harus diikuti, tahapan-tahapan penelitian ini merupakan gambaran dari langkah-langkah yang harus dilalui agar penelitian ini bisa berjalan dengan baik. Tahapan-tahapan penelitian yang harus diikuti bisa dilihat pada gambar 1:



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 3.1 di atas, berikut adalah penjabaran urutan–urutan langkah kerja sebagaimana berikut :

1. Identifikasi Masalah

Masalah yang diidentifikasi dalam penelitian ini adalah untuk melihat tingkat kepedulian masyarakat kota Pekanbaru terhadap bencana kebakaran dengan metode K-Means dan K-Medoids dengan tujuan agar penelitian ini dapat membantu Dinas Pemadam Kebakaran dan Penyelamatan Kota Pekanbaru dalam melihat tingkat kepedulian masyarakat kota pekanbaru terhadap bencana kebakaran. Dan juga memberikan solusi kepada Dinas Pemadam Kebakaran dan Penyelamatan Kota Pekanbaru dalam target pemberian dan pelatihan edukasi tentang bencana kebakaran.

2. Mempelajari Literatur

Dalam mendukung penulisan laporan skripsi ini maka penulis perlu mempelajari literatur-literatur yang akan dipakai sebagai bahan referensi dalam penelitian ini. Adapun literatur yang dipakai adalah dari jurnal-jurnal ilmiah, modul pembelajaran dan buku tentang *Data Mining* khususnya yang berkaitan dengan Algoritma K-Means dan K-Medoids dan penerapannya pada Google Colaboratory. Literatur-literatur ini akan menjadi pedoman untuk melakukan penelitian agar membantu dan memudahkan proses penelitian.

3. Mengumpulkan Data

Metode pengumpulan data dilakukan dengan melalui kuesioner skala likert yang disebarakan ke masyarakat kota Pekanbaru melalui google form.

4. Menyiapkan Dataset (*Preporcessing*)

Data kuesioner yang telah dikumpulkan sebelumnya perlu dilakukan praproses data sehingga data mentah menjadi data yang siap untuk diolah (Dataset). Pembersihan data dari nilai yang hilang atau tidak valid. Normalisasi atau standarisasi data jika diperlukan untuk memastikan semua fitur memiliki skala yang serupa.

5. Menentukan nilai k dengan metode Elbow

Proses menggunakan metode Elbow untuk menentukan jumlah kluster yang optimal memungkinkan peneliti untuk mengambil keputusan yang lebih baik terkait jumlah kluster yang

akan digunakan dalam analisis, sehingga memaksimalkan interpretabilitas dan efektivitas analisis klasterisasi.

6. Menerapkan dataset dengan algoritma K-Means dan K-Medoids
Terapkan algoritma K-Means dan K-Medoids pada data yang telah dipersiapkan. Mengikuti tahapan yang telah dijelaskan pada landasan teori.
7. Perbandingan metode
Perbandingan metode dilakukan dengan menampilkan jumlah *cluster* pada kedua metode dengan Davies Bouldin Index (DBI). Nilai Davies Bouldin-Index menggambarkan seberapa baik cluster yang dibentuk. Semakin kecil nilai Davies-Bouldin Index atau semakin mendekati nilai 0 menunjukkan seberapa baik klaster yang diperoleh.
8. Menarik kesimpulan dari hasil
Tarik kesimpulan dari kluster yang dihasilkan dan interpretasikan makna dari setiap kluster yang terbentuk

3. Hasil dan Pembahasan

Analisis data pada penelitian ini menggunakan metode K-Means dan K-Medoids untuk melakukan pengelompokkan tingkat kepedulian masyarakat terhadap bencana kebakaran yang ada di kota Pekanbaru. Untuk penentuan jumlah cluster optimal digunakan metode Elbow dengan K-Means dan K-Medoids lalu di pilih salah satu. Berdasarkan data yang didapat yaitu data kuesioner yang sudah diisikan oleh masyarakat kota Pekanbaru melalui Google Form. Adapun datanya adalah sebagai berikut:

Tabel 1. Data Kuesioner

timestamp	Nama Lengkap	Usia	Jenis Kelamin	Pekerjaan	Tempat Tinggal Saat Ini (Kecamatan) :	Q1.1	Q1.2	Q1.3	Q1.4	Q1.5	Q2.1	Q2.2	Q2.3	Q2.4	Q2.5	Q3.1	Q3.2	Q3.3	Q3.4	Q3.5	Q4.1	Q4.2	Q4.3	Q4.4	Q4.5	Q5.1	Q5.2	Q5.3	Q5.4	Q5.5
1/16/2024 17:24:32	Shel....	22	P	Pelajar/Mahasiswa	Rumbai	4	4	5	5	4	3	3	3	3	2	4	4	2	5	5	4	4	4	4	2	4	5	4	5	5
1/16/2024 17:34:25	Mhd.....	23	L	Pelajar/Mahasiswa	Rumbai	3	4	3	2	3	1	1	2	1	2	2	4	1	1	1	1	2	2	2	1	1	2	2	2	2
1/16/2024 17:34:53	Rud.....	21	L	Pelajar/Mahasiswa	Rumbai	3	4	4	5	3	1	4	5	4	4	1	5	4	1	1	1	5	1	3	1	3	3	5	5	5
...
5/16/2024 16:51:27	Hus.....	23	P	Lainnya	Senapelan	1	2	1	5	2	2	5	2	4	1	3	2	2	3	1	2	1	3	2	1	4	3	2	1	5
6/6/2024 16:54:40	Muh.....	35	L	Tidak/Belum Bekerja	Senapelan	2	3	1	2	4	3	2	4	2	2	3	5	2	4	2	3	4	5	2	4	3	4	1	2	4
6/6/2024 16:56:46	Ela.....	34	P	Tidak/Belum Bekerja	Sail	4	2	2	5	2	3	3	2	2	5	3	2	4	1	4	5	3	2	4	4	5	4	2	1	2

Merupakan data kuesioner yang didapatkan melalui media Google Form terhitung dari tanggal 16 Januari hingga 06 Juni 2024 dengan total 212 responden. Akan tetapi data ini perlu melalui tahap praproses data terlebih dahulu sebelum dapat dilakukan perhitungan dengan menggunakan metode K-Means dan K-Medoids dengan *tools* Google Colaboratory.

Data kuesioner perlu dirata-ratakan terlebih dahulu untuk memberikan gambaran secara umum bentuk tingkat kepedulian masyarakat dalam bencana kebakaran adapaun datanya adalah sebagai berikut

Data pada gambar 2 merupakan data kuesioner yang didapatkan melalui media Google Form terhitung dari tanggal 16 Januari hingga 06 Juni 2024 dengan total 212 responden. Akan tetapi data ini perlu melalui tahap praproses data terlebih dahulu sebelum dapat dilakukan perhitungan dengan menggunakan metode K-Means dan K-Medoids dengan *tools* Google Colaborator.

Tabel 2. Rata-rata Nilai Kuesioner

Kode	Avg_Q1	Avg_Q2	Avg_Q3	Avg_Q4	Avg_Q5	Kepedulian Secara Keseluruhan
R001	4,4	2,8	4	3,6	4,6	3,88
R002	3	1,4	1,8	1,6	1,8	1,92
R003	3,8	3,6	2,4	2,2	4,2	3,24
..
R210	2,2	2,8	2,2	1,8	3	2,4
R211	2,4	2,6	3,2	3,6	2,8	2,92
R212	3	3	2,8	3,6	2,8	3,04

Selanjutnya dilakukan analisis deskriptif untuk menjelaskan dan meringkas karakteristik dari kumpulan data yang ada. Tujuan utama dari analisis deskriptif adalah memberikan gambaran yang jelas dan ringkas mengenai data yang ada, sehingga dapat memberikan wawasan awal sebelum dilakukan analisis lebih mendalam. Adapun beberapa data yang akan dijelaskan adalah sebagai berikut:

1. Frekuensi Rata-rata Tingkat Kepedulian Responden

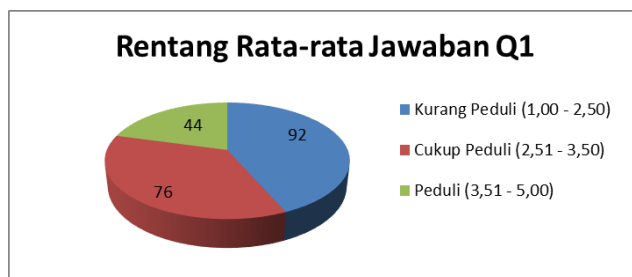
Berdasarkan data dari Tabel 3, berikut ini adalah frekuensi rata-rata tingkat kepedulian masyarakat terhadap bencana kebakaran :

a. Rata-rata Jawaban dari Parameter Q1

Tabel 3. Rata-rata Jawaban Responden Parameter Q1

Kategori	Rentang Rata-rata Jawaban Q1	Total
1	Kurang Peduli (1,00 - 2,50)	92
2	Cukup Peduli (2,51 - 3,50)	76
3	Peduli (3,51 - 5,00)	44

Parameter Q1 merupakan kumpulan pertanyaan untuk Pemahaman dan Kesadaran. Untuk Kategori 1 (Kurang Peduli) sebanyak 92 responden, untuk Kategori 2 (Cukup Peduli) sebanyak 76 responden, dan untuk Kategori 3 (Peduli) sebanyak 44 responden. Adapun tampilan diagram lingkarannya adalah sebagai berikut :



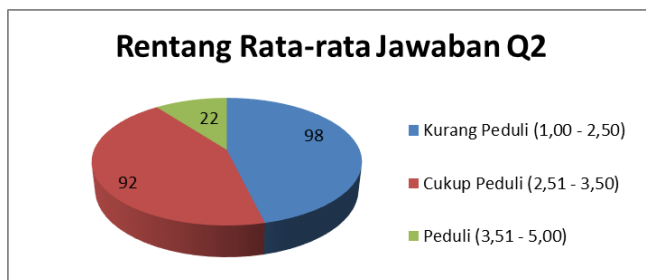
Gambar 2. Diagram Lingkaran Rentang Rata-rata Jawaban Q1

b. Rata-rata Jawaban dari Parameter Q2

Tabel 4. Rata-rata Jawaban Responden Parameter Q2

Kategori	Rentang Rata-rata Jawaban Q2	Total
1	Kurang Peduli (1,00 - 2,50)	98
2	Cukup Peduli (2,51 - 3,50)	92
3	Peduli (3,51 - 5,00)	22

Parameter Q2 merupakan kumpulan pertanyaan untuk Partisipasi dalam Pencegahan dan Respons Terhadap Kebakaran. Untuk Kategori 1 (Kurang Peduli) sebanyak 98 responden, untuk Kategori 2 (Cukup Peduli) sebanyak 92 responden, dan untuk Kategori 3 (Peduli) sebanyak 22 responden. Adapun tampilan diagram lingkarannya adalah sebagai berikut:



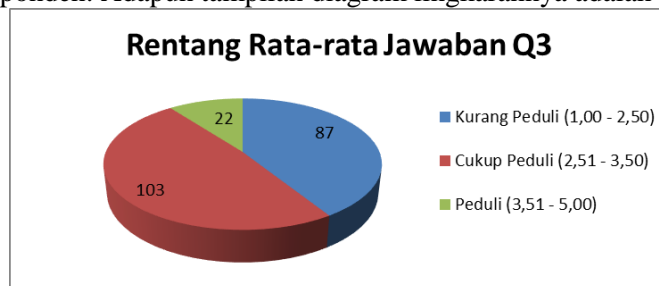
Gambar 3. Diagram Lingkaran Rentang Rata-rata Jawaban Q2

c. Rata-rata Jawaban dari Parameter Q3

Tabel 5. Rata-rata Jawaban Responden Parameter Q3

Kategori	Rentang Rata-rata Jawaban Q3	Total
1	Kurang Peduli (1,00 - 2,50)	87
2	Cukup Peduli (2,51 - 3,50)	103
3	Peduli (3,51 - 5,00)	22

Parameter Q3 merupakan kumpulan pertanyaan untuk Partisipasi dalam Kampanye Kesadaran Kebakaran. Untuk Kategori 1 (Kurang Peduli) sebanyak 87 responden, untuk Kategori 2 (Cukup Peduli) sebanyak 103 responden, dan untuk Kategori 3 (Peduli) sebanyak 22 responden. Adapun tampilan diagram lingkarannya adalah sebagai berikut:

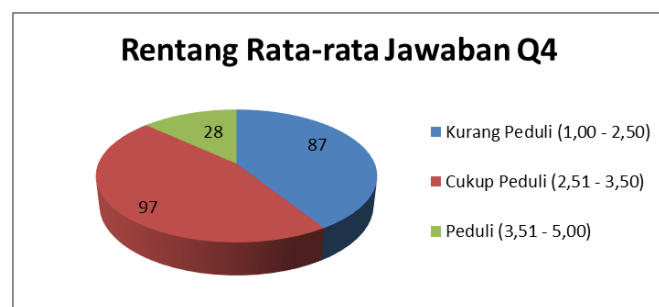
**Gambar 4.** Diagram Lingkaran Rentang Rata-rata Jawaban Q3

d. Rata-rata Jawaban dari Parameter Q4

Tabel 6. Rata-rata Jawaban Responden Parameter Q4

Kategori	Rentang Rata-rata Jawaban Q4	Total
1	Kurang Peduli (1,00 - 2,50)	87
2	Cukup Peduli (2,51 - 3,50)	97
3	Peduli (3,51 - 5,00)	28

Parameter Q4 merupakan kumpulan pertanyaan untuk Kontribusi Sosial dan Komunitas. Untuk Kategori 1 (Kurang Peduli) sebanyak 87 responden, untuk Kategori 2 (Cukup Peduli) sebanyak 97 responden, dan untuk Kategori 3 (Peduli) sebanyak 28 responden. Adapun tampilan diagram lingkarannya adalah sebagai berikut:

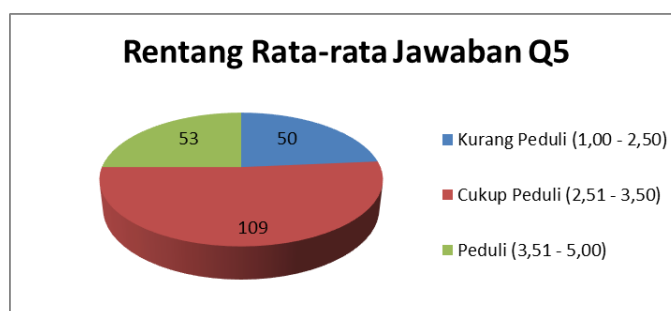
**Gambar 5.** Diagram Lingkaran Rentang Rata-rata Jawaban Q4

e. Rata-rata Jawaban dari Parameter Q5

Tabel 7. Rata-rata Jawaban Responden Parameter Q5

Kategori	Rentang Rata-rata Jawaban Q5	Total
1	Kurang Peduli (1,00 - 2,50)	50
2	Cukup Peduli (2,51 - 3,50)	109
3	Peduli (3,51 - 5,00)	53

Parameter Q5 merupakan kumpulan pertanyaan untuk Persepsi tentang Tanggung Jawab Bersama. Untuk Kategori 1 (Kurang Peduli) sebanyak 50 responden, untuk Kategori 2 (Cukup Peduli) sebanyak 109 responden, dan untuk Kategori 3 (Peduli) sebanyak 53 responden. Adapun tampilan diagram lingkarannya adalah sebagai berikut:



Gambar 6. Diagram Lingkaran Rentang Rata-rata Jawaban Q5

Pada praproses data ini dilakukan dengan mengambil data mentah dari Tabel 5.1 lalu selanjutnya dilakukan perhitungan rata-rata dari nilai jawaban pada setiap parameter Q1, Q2, Q3, Q4, dan Q5 serta rata-rata keseluruhan dari semua parameter. Dan selanjutnya ditransformasikan menjadi sebuah dataset yang baru dengan tambahan data dari kategori-kategori berikut :

Tabel 8. Kategori Data

Usia	Kategori	Kecamatan	Kategori
18-25 Tahun	1	Binawidya	1
26-35 Tahun	2	Bukit Raya	2
36-45 Tahun	3	Kulim	3
> 46 Tahun	4	Lima Puluh	4
		Marpoyan Damai	5
		Payung Sekaki	6
		Pekanbaru Kota	7
		Rumbai Barat	8
		Rumbai	9
		Rumbai Timur	10
		Sail	11
		Senapelan	12
		Sukajadi	13
		Tuah Madani	14
		Tenayan Raya	15

Jenis Kelamin	Kategori
Laki-laki	1
Perempuan	2

Pekerjaan	Kategori
Tidak/Belum Bekerja	0
Pelajar/Mahasiswa	1
Pegawai Pemerintahan	2
Pegawai Swasta	3
Wiraswasta/Usaha Pribadi	4
Lainnya	5

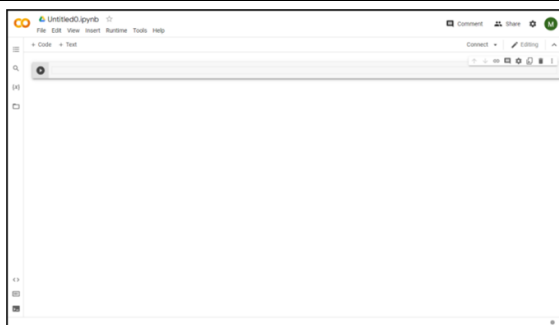
Setelah melakukan persiapan data selanjutnya melakukan pengujian dengan Tools Google Colaboratory. Adapun tahapannya adalah sebagai berikut:

1. Menyiapkan Data

Data yang disiapkan merupakan Dataset yang terdapat pada Tabel 5.13 yang disimpan dalam format .xlsx. data ini nantinya digunakan sebagai bahan perhitungan analisis yang diunggah ke halaman Google Colaboratory.

2. Implementasi Dengan Tools Google Colaboratory

Melakukan pengelompokkan data cluster data tingkat kepedulian masyarakat pada bencana kebakaran di kota Pekanbaru dengan penerapan Algoritma K-Means dan K-Medoids pada Google Colaboratory. Berikut merupakan halaman utama Google Colaboratory:



Gambar 7. Tools Google Colaboratory

Pada bagian ini akan dibahas mengenai tahapan pengujian dataset. Adapun tahapan K-Means dan K-Medoids menggunakan *Google Colaboratory* adalah sebagai berikut:

a. Unggah Data Pada *Google Drive*

Upload data dilakukan untuk memasukkan data yang telah disiapkan sebelumnya dalam bentuk format *Excel*. Selanjutnya memasukkan data ke *Google Drive* agar data dapat dipanggil dengan *coding* pada *Google Colaboratory*.

b. Mengimport Library

Pada implementasi *Google Colaboratory* memiliki kelas yaitu kelas proses. Selain itu, perangkat lunak ini menggunakan *library python*.

```
# Mengimport library
!pip install scikit-learn-extra
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from sklearn.cluster import KMeans
from sklearn_extra.cluster import KMedoids
```

Gambar 8. Import Library

Berdasarkan pada gambar diatas berikut adalah penjelasan dari masing-masing *library python* yang di masukkan pada *cell code Google Colaboratory*:

Tabel 9. Deskripsi *Import Library*

No	Kode	Deskripsi
1	<code>!pip install scikit-learn-extra</code>	Untuk menginstall library K-Medoids.
2	<code>import numpy as np</code>	Menyediakan obyek array, menganalisis numerik, melakukan matrix, dan operasi matematika.
3	<code>import matplotlib.pyplot as plt</code>	Untuk membuat plot.
4	<code>import pandas as pd</code>	Untuk menganalisis, membersihkan, mengeksplorasi.
5	<code>from sklearn.cluster import KMeans</code>	Untuk Menghitung K-Means.
6	<code>from sklearn_extra.cluster import KMedoids</code>	Untuk Menghitung K-Medoids.

Langkah pertama adalah menginstal pustaka *scikit-learn-extra* menggunakan perintah `!pip install scikit-learn-extra`. Library ini diperlukan untuk menjalankan algoritma K-Medoids. Setelah itu, beberapa Library penting diimpor ke dalam program. Library *numpy* digunakan untuk menyediakan objek array, analisis numerik, operasi matriks, serta operasi matematika lainnya. Selanjutnya, Library *matplotlib.pyplot* diimpor untuk membuat visualisasi berupa grafik atau plot. Library *pandas* digunakan untuk analisis data, termasuk membersihkan, mengeksplorasi, dan memanipulasi data. Kemudian, algoritma K-Means dapat dihitung menggunakan modul *KMeans* dari Library *sklearn cluster*. Terakhir, untuk

perhitungan algoritma K-Medoids, digunakan modul KMedoids yang tersedia di Library sklearn_extra cluster.

c. Memanggil Data Dari *Google Drive*

Pada tahap ini melakukan pemanggilan data yang sudah diunggah ke Google Drive sebelumnya, dengan kode pada cell code seperti pada gambar berikut:

```
# Mengimpor dataset
from google.colab import drive
drive.mount('/content/drive/')
dataset = pd.read_excel('/content/drive/MyDrive/dataset_kebakaran.xlsx')
X = dataset.drop(['Kode', 'Jenis_Kelamin', 'Pekerjaan', 'Kecamatan'],
axis=1)
X.head()
```

Gambar 9. Pemanggilan Dataset

Selanjutnya menjalankan cell code pada Gambar diatas dan menghasilkan pemanggilan data yang sama seperti data yang kita unggah pada Google Drive sebelumnya. Untuk data Kode, Jenis Kelamin, Pekerjaan dan Kecamatan di hilangkan karena tidak dijadikan sebagai bahan perhitungan. Adapun tampilan dari cell code diatas adalah sebagai berikut:

	Usia	Avg_Q1	Avg_Q2	Avg_Q3	Avg_Q4	Avg_Q5	Kepedulian Secara Keseluruhan
0	22	4.4	2.8	4.0	3.6	4.6	3.88
1	23	3.0	1.4	1.8	1.6	1.8	1.92
2	21	3.8	3.6	2.4	2.2	4.2	3.24
3	19	1.8	1.0	1.0	1.0	2.2	1.40
4	23	2.8	3.0	2.8	2.8	3.2	2.92

Gambar 10. Dataset

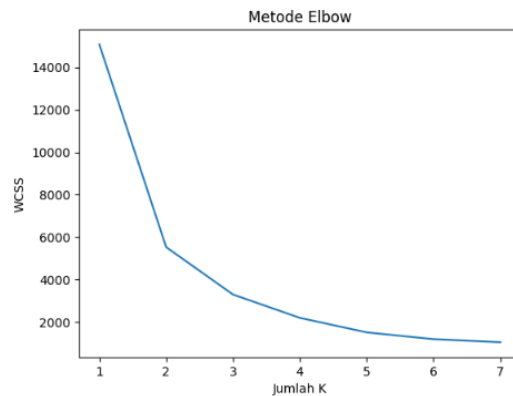
d. Optimasi Nilai K Dengan Metode *Elbow*

Menentukan nilai k optimal dilakukan dengan menerapkan Metode Elbow. Jika nilai k atau cluster sebelumnya dengan nilai k atau cluster selanjutnya menampilkan titik sudut pada gambar grafik atau nilainya menunjukkan penurunan cukup tinggi maka k atau cluster tersebut yang terbaik dengan perhitungan Within Cluster Sum of Square Error (WCSS) yang adalah perbandingan antara jumlah kuadrat dan jumlah kuadrat yang mewakili jumlah data yang dikumpulkan per centroid. Berikut adalah tampilan cell code dari coding yang dijalankan:

```
from sklearn.cluster import KMeans
wcss = []
for i in range(1, 8):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state =
None)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 8), wcss)
plt.title('Metode Elbow')
plt.xlabel('Jumlah K')
plt.ylabel('WCSS')
plt.show()
```

Gambar 11. Cell Code Optimasi Nilai K

Berikut visualisasi Metode Elbow berdasarkan perhitungan WCSS dengan range k atau cluster 1 sampai 8 dalam bentuk grafik:



Gambar 12. Optimasi K Dengan Metode Elbow

Berdasarkan dari Gambar 12 dapat diketahui pada metode *Elbow*, nilai SSE diharapkan menjadi minimal dengan mencari jumlah *cluster* yang tepat dan titik tengah *cluster* yang tepat (*centroid*). Pada grafik diatas menampilkan hasil *clustering* data tingkat kepedulian masyarakat dalam bencana kebakaran dengan $k=3$ dan membentuk sudut siku-siku serta pengurangan nilai WCSS tidak terlalu signifikan seperti pada gambar berikut:

Cluster	WCSS	Selisih
1	15079.823955	NaN
2	5523.804127	-9556.019828
3	3303.817842	-2219.986285
4	2189.952052	-1113.865790
5	1519.819309	-670.132743
6	1186.226428	-333.592881
7	1071.138002	-115.088426

Gambar 13. Selisih Nilai WCSS

Pada Gambar 13. dapat diketahui nilai WCSS untuk Cluster 1 berada pada 15079,82 yang memiliki selisih sebesar 9556,01 terhadap Cluster 2 yang memiliki nilai WCSS sebesar 5523,80, dan seterusnya pada nilai WCSS untuk Cluster 2 memiliki selisih sebesar 2219,98 terhadap Cluster 3 yang memiliki nilai WCSS sebesar 3303,81 sampai seterusnya sampai pada Cluster 7. Yang perlu diperhatikan adalah apabila penurunan nilai WCSS sudah tidak terlalu signifikan maka jumlah cluster sudah dapat dikatakan optimal sehingga dapat dikatakan nilai $k=3$

e. Menghitung K-Means Pada Dataset

Setelah menghitung nilai k optimal selanjutnya data diuji menggunakan algoritma K-Means berdasarkan nilai $k=3$. Adapun tampilan cellcode-nya adalah sebagai berikut :

```
kmeans = KMeans(n_clusters = 3, init = 'k-means++', random_state = None)
y_kmeans = kmeans.fit_predict(X)

dataset['Cluster_KMeans'] = y_kmeans
dataset_kmeans = dataset
dataset_kmeans
```

Gambar 14. Cellcode Perhitungan Metode K-Means

Pada perhitungannya didapat 3 kelompok data *cluster* responden dengan *DataFrame Cluster* dan *Cluster Center* serta Jumlah Anggota tiap *Cluster* seperti pada gambar-gambar berikut:

	Kode	Usia	Jenis_Kelamin	Pekerjaan	Kecamatan	Avg_Q1	Avg_Q2	Avg_Q3	Avg_Q4	Avg_Q5	Kepedulian Secara Keseluruhan	Cluster_MMeans	
0	R001	22		2	1	9	4.4	2.8	4.0	3.6	4.6	3.88	1
1	R002	23		1	1	9	3.0	1.4	1.8	1.6	1.8	1.92	1
2	R003	21		1	1	9	3.8	3.6	2.4	2.2	4.2	3.24	1
3	R004	19		1	5	9	1.8	1.0	1.0	1.0	2.2	1.40	1
4	R005	23		1	1	9	2.8	3.0	2.8	2.8	3.2	2.92	1
...
207	R208	25		2	5	11	2.8	3.6	2.6	2.8	2.4	2.84	1
208	R209	29		2	0	12	2.2	3.0	2.2	2.2	2.8	2.48	1
209	R210	23		2	5	12	2.2	2.8	2.2	1.8	3.0	2.40	1
210	R211	35		1	0	12	2.4	2.6	3.2	3.6	2.8	2.92	0
211	R212	34		2	0	11	3.0	3.0	2.8	3.6	2.8	3.04	0

212 rows x 12 columns

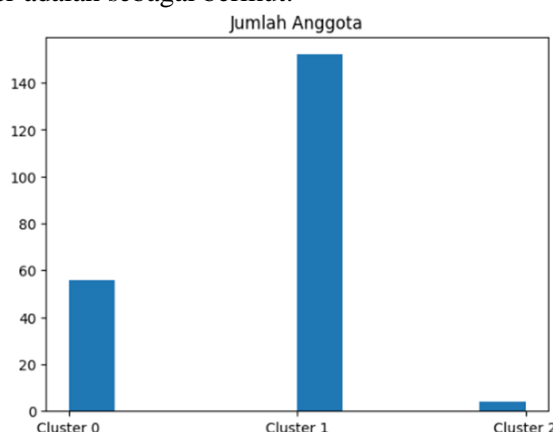
Gambar 15. Data Frame Cluster K-Means pada Dataset

Data pada gambar 14 merupakan merupakan *output* dari *cluster* pada setiap responden. Untuk *centroid* akhir dapat dilihat pada gambar 5.18 berikut :

	Usia	Q1	Q2	Q3	Q4	Q5	Kepedulian Secara Keseluruhan
0	35.964286	2.478571	2.525000	2.578571	2.660714	2.900000	2.628571
1	22.769737	2.843421	2.584211	2.659211	2.668421	3.086842	2.768421
2	60.750000	3.450000	2.450000	2.250000	1.900000	2.550000	2.520000

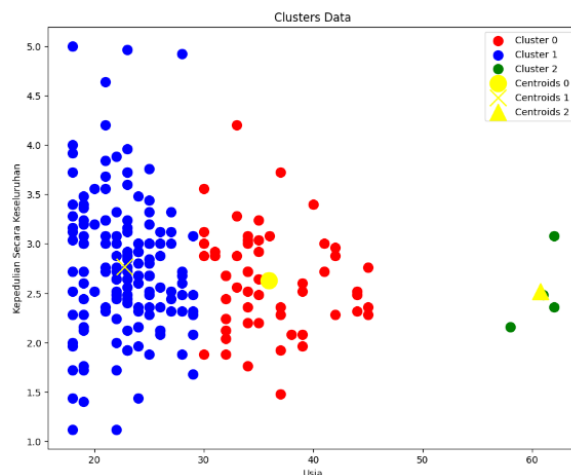
Gambar 16. Cluster Center Dataset Metode K-Means

Untuk dataset Cluster 0 memiliki nilai tengah cluster Usia sebesar 35,96, untuk Parameter Q1 sebesar 2,47, untuk Parameter Q2 sebesar 2,52, untuk Parameter Q3 sebesar 2,57, untuk Parameter Q4 sebesar 2,66, untuk Parameter Q5 sebesar 2,90 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,62. Lalu untuk dataset Cluster 1 memiliki nilai tengah cluster Usia sebesar 22,76, untuk Parameter Q1 sebesar 2,84, untuk Parameter Q2 sebesar 2,58, untuk Parameter Q3 sebesar 2,65, untuk Parameter Q4 sebesar 2,66, untuk Parameter Q5 sebesar 3,08 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,76. Dan terakhir untuk dataset Cluster 2 memiliki nilai tengah cluster Usia sebesar 60,75, untuk Parameter Q1 sebesar 3,45, untuk Parameter Q2 sebesar 2,45, untuk Parameter Q3 sebesar 2,25, untuk Parameter Q4 sebesar 1,90, untuk Parameter Q5 sebesar 2,55 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,52. Adapun tampilan grafik jumlah anggota tiap cluster adalah sebagai berikut:



Gambar 17. Grafik Jumlah Anggota Tiap Cluster Metode K-Means

Terdapat sebanyak 56 anggota pada Cluster 0, pada Cluster 1 sebanyak anggota 152, dan untuk Cluster 2 sebanyak anggota 4. Cluster yang telah terbentuk selanjutnya dilakukan visualisasi dalam bentuk grafik *scatterplot* seperti pada gambar dibawah ini:



Gambar 18. Visualisasi Cluster Metode K-Means

Pada gambar diatas terdapat bulatan besar dan kecil, yang mana tanda besar berwarna kuning merupakan cluster center dari Cluster 0, Cluster 1 dan Cluster 2, warna merah merupakan data yang terbentuk sebagai Cluster 0 sementara warna biru merupakan data yang terbentuk sebagai Cluster 1 dan yang terakhir Cluster 2 berwarna hijau.

f. Menghitung K-Medoids Pada Dataset

Selanjutnya melakukan perhitungan menggunakan algoritma K-Medoids berdasarkan nilai $k=3$. Adapun tampilan cellcode-nya adalah sebagai berikut :

```
kmedoids = KMedoids(n_clusters = 3, init = 'k-medoids++', random_state = 30)
y_kmedoids = kmedoids.fit_predict(X)

dataset['Cluster_KMedoids'] = y_kmedoids
dataset_kmedoids = dataset
dataset_kmedoids
```

Gambar 19. Cellcode Perhitungan Metode K-Medoids

Pada perhitungannya didapat 3 kelompok data cluster responden dengan DataFrame Cluster dan Cluster Center serta Jumlah Anggota tiap Cluster seperti pada gambar-gambar berikut :

	Kode	Usia	Jenis_Kelamin	Pekerjaan	Kecamatan	Avg_Q1	Avg_Q2	Avg_Q3	Avg_Q4	Avg_Q5	Kepedulian Secara Keseluruhan	Cluster_KMedoids
0	R001	22	2	1	9	4.4	2.8	4.0	3.6	4.6	3.88	0
1	R002	23	1	1	9	3.0	1.4	1.8	1.6	1.8	1.92	0
2	R003	21	1	1	9	3.8	3.6	2.4	2.2	4.2	3.24	0
3	R004	19	1	5	9	1.8	1.0	1.0	1.0	2.2	1.40	0
4	R005	23	1	1	9	2.8	3.0	2.8	2.8	3.2	2.92	0
...
207	R208	25	2	5	11	2.8	3.6	2.6	2.8	2.4	2.84	0
208	R209	29	2	0	12	2.2	3.0	2.2	2.2	2.8	2.48	1
209	R210	23	2	5	12	2.2	2.8	2.2	1.8	3.0	2.40	0
210	R211	35	1	0	12	2.4	2.6	3.2	3.6	2.8	2.92	1
211	R212	34	2	0	11	3.0	3.0	2.8	3.6	2.8	3.04	1

212 rows x 12 columns

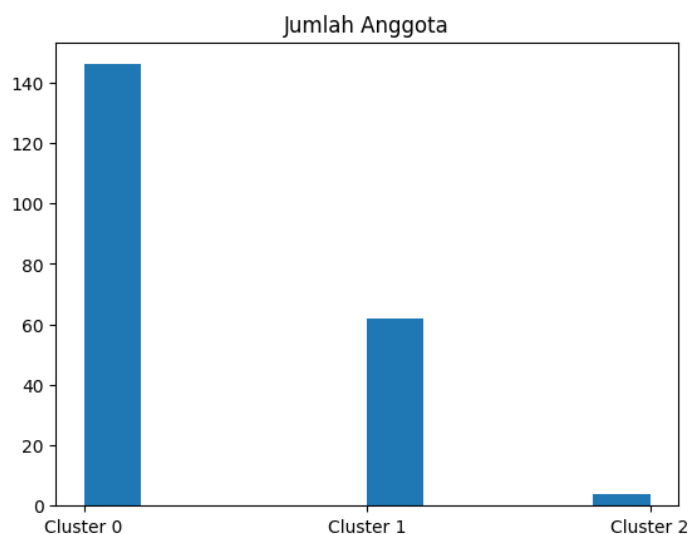
Gambar 20. DataFrame Cluster K- Medoids pada Dataset

Data pada gambar 5.17 merupakan merupakan output dari cluster pada setiap responden. Untuk centroid akhir dapat dilihat pada gambar 5.18 berikut :

	Usia	Q1	Q2	Q3	Q4	Q5	Kepedulian Secara Keseluruhan
0	23.0	3.0	2.6	3.0	2.6	3.2	2.88
1	34.0	2.4	3.0	2.8	2.8	2.6	2.72
2	61.0	3.0	2.8	3.0	1.6	2.0	2.48

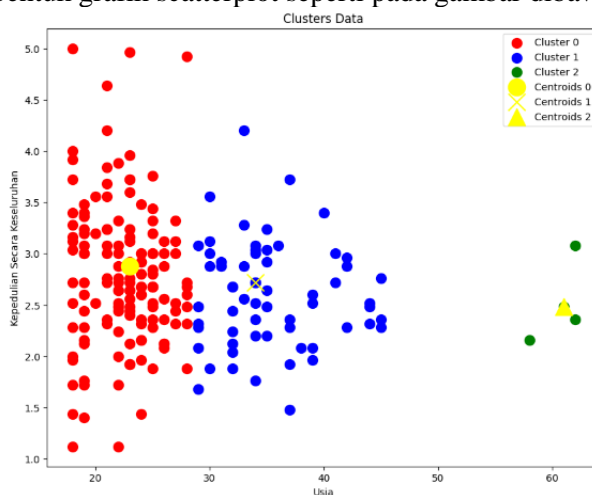
Gambar 21. Cluster Center Dataset Metode K- Medoids

Untuk dataset Cluster 0 memiliki nilai tengah cluster Usia sebesar 23, untuk Parameter Q1 sebesar 3,0, untuk Parameter Q2 sebesar 2,6, untuk Parameter Q3 sebesar 3,0, untuk Parameter Q4 sebesar 2,6, untuk Parameter Q5 sebesar 3,5 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,88. Lalu untuk dataset Cluster 1 memiliki nilai tengah cluster Usia sebesar 34, untuk Parameter Q1 sebesar 2,4, untuk Parameter Q2 sebesar 3,0, untuk Parameter Q3 sebesar 2,8, untuk Parameter Q4 sebesar 2,8, untuk Parameter Q5 sebesar 2,6 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,72. Dan terakhir untuk dataset Cluster 2 memiliki nilai tengah cluster Usia sebesar 61, untuk Parameter Q1 sebesar 3,0, untuk Parameter Q2 sebesar 2,8, untuk Parameter Q3 sebesar 3,0, untuk Parameter Q4 sebesar 1,6, untuk Parameter Q5 sebesar 2,0 dan untuk Parameter Kepedulian Secara Keseluruhan sebesar 2,48. Adapun tampilan grafik jumlah anggota tiap cluster adalah sebagai berikut :



Gambar 22. Grafik Jumlah Anggota Tiap Cluster Metode K- Medoids

Terdapat sebanyak 146 anggota pada Cluster 0, pada Cluster 1 sebanyak anggota 62, dan untuk Cluster 2 sebanyak anggota 4. Cluster yang telah terbentuk selanjutnya dilakukan visualisasi dalam bentuk grafik scatterplot seperti pada gambar dibawah ini :



Gambar 23. Visualisasi Cluster Metode K- Medoids

Pada gambar diatas terdapat bulatan besar dan kecil, yang mana bulatan besar berwarna kuning merupakan cluster center dari Cluster 0, Cluster 1 dan Cluster 2, warna merah merupakan data yang terbentuk sebagai Cluster 0 sementara warna biru merupakan data yang terbentuk sebagai Cluster 1 dan yang terakhir Cluster 2 berwarna hijau.

4. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Means dan K-Medoids menghasilkan kluster yang berbeda dalam hal usia dan tingkat kepedulian masyarakat terhadap bencana kebakaran. Metode K-Means membentuk tiga kluster dengan rata-rata usia yang lebih tinggi pada Cluster 0 dan Cluster 2, serta menunjukkan variasi yang lebih besar dalam data kepedulian secara keseluruhan. Sebaliknya, metode K-Medoids menghasilkan kluster dengan usia rata-rata yang lebih rendah dan lebih stabil, terutama di Cluster 2. K-Medoids juga menunjukkan nilai kepedulian yang lebih tinggi di Cluster 0, sementara K-Means mencatatkan nilai kepedulian yang lebih tinggi di Cluster 1 dan Cluster 2. Secara keseluruhan, K-Means lebih mencerminkan variasi dalam data, sementara K-Medoids memberikan hasil yang lebih stabil dan kurang terpengaruh oleh data ekstrem.

Daftar Pustaka

- [1] Athifaturrofifah, Goejantoro, R., & Yuniarti, D. (2019). 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). *Jurnal EKSPONENSIAL*.
- [2] Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI: Jurnal Teknologi Informasi*.
- [3] Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi (JATI)*.
- [4] Hamzah, R. A. N., Widiyasosno, N., & Kurniati, N. I. (2019). Implementasi Location Based Service pada Aplikasi Pannic Button Berbasis Android. *SAIS - Scientific Articles of Informatics Students*.
- [5] Handayani, F. (2022). Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar. *Jurnal Teknologi Dan Informasi (JATI)*.
- [6] Luchia, N. T., Handayani, H., Hamdi, F. S., Erlangga, D., & Octavia, S. F. (2022). Perbandingan K-Means dan K-Medoids Pada Pengelompokan Data Miskin di Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*.
- [7] Mariam, S., Handayani, F., & Jualiane, C. (2023). Penerapan Algoritma Clustering K-Means Untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam. *Jurnal Teknik Informatika Dan Sistem Informasi*.
- [8] Nawangsih, I., Puspita, R., & Suherman. (2021). Implementasi Algoritma K-Means Dalam Mengkategorikan Produk Terlaris Dan Kurang Laris Pada Toko Alfamart Cikarang. *PELITA TEKNOLOGI*.
- [9] Palembang, C. F., Matdoan, M. Y., & Palembang, S. P. (2022). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokan Tingkat Kebahagiaan Provinsi Di Indonesia. *BULLET □: Jurnal Multidisiplin Ilmu*.
- [10] Parlambang, B., & Fauziah. (2020). Implementasi Algoritma K-Means Dalam Proses Penilaian Kuesioner Kepada Dosen Guna Mendukung Kepuasan Mahasiswa Terhadap Dosen. *Jurnal Ilmiah Teknologi Dan Rekayasa*.
- [11] Rahma, L., Syaputra, H., Mirza, A. H., & Purnamasari, S. D. (2021). Objek Deteksi Makanan Khas Palembang Menggunakan Algoritma YOLO (You Only Look Once). *Jurnal Nasional Ilmu Komputer*.
- [12] Siburian, T., Safii, M., & Parlina, I. (2019). Penerapan Algoritma K-Means Clustering untuk Pengelompokan Harga Eceran Beras di Pasar Tradisional Berdasarkan Wilayah Kota. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*.
- [13] Sinaga, S. M., Hardinata, J. T., & Fauzan, M. (2021). Implementasi Data Mining Clustering Tingkat Kepuasan Konsumen Terhadap Pelayanan Go-Jek. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*.
- [14] Sipayung, A. T., Saifullah, & Winanjaya, R. (2020). Penerapan Metode K-Means Dalam Mengelompokkan Banyaknya Desa / Kelurahan Menurut Jenis Pencemaran Lingkungan Hidup Berdasarkan Provinsi. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*.

- [15] Wahyudi, I., Sulthan, M. B., & Suhartini, L. (2021). Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan. Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM).



ZONasi: Jurnal Sistem Informasi

Is licensed under a [Creative Commons Attribution International \(CC BY-SA 4.0\)](https://creativecommons.org/licenses/by-sa/4.0/)