

# Penerapan Algoritma K-Medoids dalam Menentukan Daerah Rawan Banjir di Kabupaten Karawang

Cepy Sukmayadi\*, Aji Primajaya\*\*, Iqbal Maulana\*\*\*

\* Teknik Informatika, Ilmu Komputer, Universitas Singaperbangsa Karawang

\*\* Teknik Informatika, Ilmu Komputer, Universitas Singaperbangsa Karawang

\*\*\* Teknik Informatika, Ilmu Komputer, Universitas Singaperbangsa Karawang

\*cepy.sukmayadi17009@student.unsika.ac.id, \*\*aji.primajaya@staff.unsika.ac.id, \*\*\*Iqbal.maulana@staff.unsika.ac.id

## ABSTRACT

Flood disasters often occur during the rainy season. Karawang is one area that is often flooded. Based on the risk index from BNPB, the flood disaster in Karawang affected 84% of the community, so efforts need to be made to reduce and overcome flood disasters. These problems are the beginning of efforts that need to be known which areas are prone to flooding. Therefore, this study aims to determine flood-prone areas in Karawang as an initial effort in tackling flood disasters. The research was conducted by classifying flood-prone areas using the k-medoids algorithm. K-Medoids uses the partition clustering method to group lists and objects into a number of clusters. This algorithm uses objects in a collection of objects that represent a cluster. The attributes used are flood-causing factors such as rainfall, elevation (soil height), population density, and distance to the river. The results of the study found three potential floods, namely low, medium, and high. There are 1 sub-district with low flood potential, 24 sub-districts with moderate flood potential, and 5 sub-districts with high flood potential. The test results using the silhouette coefficient get a value of 0.370.

**Keyword:** Flood, Clustering, Data Mining, K-Medoids.

## 1. Introduction

Bencana banjir merupakan bencana alam yang seringkali terjadi ketika musim penghujan tiba. Musim hujan ini berlangsung antara bulan Oktober sampai Maret. Berdasarkan prakiraan BMKG puncak musim hujan terjadi pada bulan Januari dan Februari. Pada bulan tersebut curah hujan yang turun begitu tinggi sehingga dapat menyebabkan aliran air meluap dan menimbulkan banjir. Daerah yang rawan terhadap banjir akan mendapatkan dampaknya [1].

Salah satu daerah yang rawan terhadap banjir adalah kabupaten Karawang. Secara topografi sebagian besar wilayah kabupaten Karawang berupa dataran rendah dengan ketinggian  $\pm 25$  mdpl. Pada bagian utara terdapat dataran pantai yang luas sedangkan pada bagian selatan merupakan wilayah yang berbukit-bukit. Kabupaten Karawang dilalui oleh sungai Citarum dan sungai Cilamaya yang mengarah ke utara Laut Jawa. Sungai Citarum juga merupakan batas alam pemisah dengan Kabupaten Bekasi sedangkan sungai Cilamaya merupakan batas alam pemisah dengan Kabupaten Subang [2].



Gambar 1. Indeks Risiko Bencana Kabupaten Karawang  
(Sumber: <https://inarisk.bnpb.go.id/>)

Berdasarkan Indeks Risiko Bencana Badan Nasional Penanggulangan Bencana, bencana banjir merupakan bencana terbesar kedua di Kabupaten Karawang setelah bencana kekeringan. Bencana ini mengakibatkan sekitar 84% jiwa terpapar dampak banjir. Dampak dari banjir dapat menyebabkan terendahnya pemukiman warga, selain itu juga dapat merendam infrastruktur, sarana ibadah, sarana pendidikan, area sawah, tambak, dan lain-lainnya.

Kerugian yang didapat dari bencana banjir tentunya harus diminimalisir. Salah satu langkah yang dapat dilakukan adalah dengan menentukan daerah yang rawan terhadap banjir sebagai upaya awal dalam

penanggulangan bencana banjir. Nantinya, dapat dijadikan acuan dalam melakukan upaya penanggulangan banjir lainnya. Oleh karena itu, pada penelitian ini dilakukan proses data mining untuk mengetahui daerah rawan banjir berdasarkan faktor penyebabnya. Data mining dilakukan dengan metode clustering untuk mengelompokkan data menggunakan algoritma *k-medoids*. Menurut Riyanto dalam [3] *k-medoids* hadir untuk mengatasi kelemahan *k-means* yang sensitif terhadap outlier karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data.

Pada penelitian sebelumnya yang dilakukan oleh Pramesti D. F, dkk (2017) dengan judul “Implementasi Metode K-Medoids untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)”. Atribut yang digunakan, yaitu *latitude*, *longitude*, *brightness*, *frp* (*fire radiative power*), dan *confidence*. Adapun hasilnya didapatkan 2 cluster dengan nilai *silhouette coefficient* sebesar 0.56745 [4]. Selanjutnya pada penelitian yang dilakukan oleh Simamora D.A.S, dkk (2017) dengan judul “Clustering Data Kejadian Tsunami Yang Disebabkan Oleh Gempa Bumi Dengan Menggunakan Algoritma K-Medoids” Atribut yang digunakan, yaitu kekuatan gempa, kedalaman pusat gempa (*focal depth*), *latitude*, *longitude*. Adapun hasilnya didapatkan 2 cluster dengan nilai *silhouette coefficient* sebesar 0,40066. [5] Penelitian lainnya yang dilakukan oleh Khusaeri dkk (2017) dengan judul “Algoritma C4.5 untuk Pemodelan Daerah Rawan Banjir Studi Kasus Kabupaten Karawang Jawa Barat”. Atribut yang digunakan, yaitu curah hujan, ketinggian wilayah, jarak ke sungai, kepadatan penduduk, dan data banjir sebagai label. Adapun hasil evaluasi menghasilkan tingkat akurasi sebesar 84,385% [6].

Pada penelitian sekarang akan dilakukan pengelompokan daerah rawan banjir di kabupaten Karawang menggunakan algoritma *k-Medoids*. Adapun, data yang digunakan merupakan faktor penyebab banjir berdasarkan penelitian sebelumnya, yaitu curah hujan, elevasi (ketinggian tanah), kepadatan penduduk, dan jarak ke sungai.

### 1.1 Missing Value

*Missing value* merupakan masalah dalam menjaga suatu kualitas data. *Missing value* dapat menyebabkan tingkat keakuratan suatu data menjadi berkurang dan menurunnya kualitas data pada saat akan dilakukan pengolahan data. Oleh karena itu, diperlukan penanganan khusus untuk mengatasi *missing value* ini. Beberapa cara yang dapat dilakukan dengan menghapus keseluruhan baris atau menghitung nilai penggantinya menggunakan mean, modus, dan lain-lain [7].

### 1.2. Min-Max Normalization

Normalisasi dalam data mining merupakan proses penskalaan nilai atribut dari data menjadi skala tertentu. Normalisasi ini digunakan untuk menyamakan skala atribut data kedalam skala yang lebih kecil seperti -1 sampai 1 atau 0 sampai 1 [4]. Salah satu metode normalisasi adalah *min-max normalization* yang merupakan metode normalisasi dengan melakukan transformasi linear pada atribut data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses [8].

Berikut ini persamaan dari *min-max normalization* :

$$v' = \frac{v - \min_A}{\max_A - \min_A} (\text{new\_max}_A - \text{new\_min}_A) + \text{new\_min}_A \quad (1)$$

Keterangan :

- v' = data setelah dinormalisasi
- v = data yang akan dinormalisasi
- minA = nilai minimal dari data yang akan dinormalisasi
- maxA = nilai maksimal dari data yang akan dinormalisasi
- new\_minA = nilai minimal setelah dinormalisasi
- new\_maxA = nilai maksimal setelah dinormalisasi

### 1.3. K-Medoids

*K-Medoids* atau *Partitioning Around Method* (PAM) adalah metode *cluster* non hirarki yang merupakan varian dari metode *k-means*. *k-medoids* hadir untuk mengatasi kelemahan *k-means* yang sensitif terhadap outlier karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [9]. *K-medoids* menggunakan metode pengelompokan partisi untuk mengelompokkan sekumpulan n objek menjadi sejumlah k *cluster*. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah *cluster*. Objek yang mewakili sebuah *cluster* disebut dengan *medoids*. *Medoid* merupakan objek yang letaknya terpusat di dalam suatu cluster sehingga baik terhadap outlier. *Cluster* dibangun dengan menghitung kedekatan yang dimiliki antara medoids dengan objek *non medoids* [10].

Langkah-langkah dari algoritma *k-medoids* adalah sebagai berikut [11]:

1. Menentukan jumlah  $k$  (*cluster*)
2. Menentukan pusat *cluster* sebanyak  $k$
3. Menghitung jarak masing-masing objek ke pusat *cluster* dan hitung totalnya
4. Menentukan pusat *cluster* baru.
5. Menghitung jarak masing-masing objek ke pusat *cluster* dan hitung totalnya
6. Hitung total simpangan ( $S$ ) dengan menghitung total *distance* baru – total *distance* lama. Jika  $S < 0$  maka tukar objek dengan data *cluster non-medoids*.
7. Ulangi Langkah 4 – 6 hingga tidak terjadi perubahan pada anggota *medoids*.

#### 1.4. Silhouette Coefficient

*Silhouette Coefficient* merupakan metode yang digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Metode ini menggabungkan nilai kohesi dan separasi. Kohesi digunakan untuk mengukur sedekat apa relasi antar satu objek dengan objek lainnya pada sebuah *cluster* dan separasi digunakan untuk menghitung seberapa jauh sebuah *cluster* berpisah dengan cluster yang lain [12]. Nilai *silhouette coefficient* rata-rata dari setiap data atau objek pada sebuah cluster adalah merupakan suatu takaran yang menunjukkan seketat apada objek-objek dimasukkan pada suatu cluster tertentu [13]

Berikut ini tahapan-tahapan perhitungan *silhouette coefficient* :

1. Hitung rata-rata dari suatu data misalkan  $i$  dengan semua dokumen lain yang berada dalam satu *cluster*. Berikut ini persamaanya :

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

Dimana  $a(i)$  adalah perbedaan rata-rata objek ( $i$ ) ke semua objek pada satu cluster  $A$ ,  $|A|$  adalah jumlah data pada cluster  $A$ , dan  $d(i, j)$  adalah jarak antara data  $i$  dan  $j$ .

2. Hitung rata-rata jarak data  $i$  tersebut dengan semua data di *cluster* lain dan diambil nilai terkecilnya. Berikut ini persamaanya :

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

Dimana  $d(i, C)$  adalah jarak rata-rata data  $i$  dengan semua objek pada *cluster* lain  $C$  dimana  $A \neq C$ ,  $|C|$  adalah jumlah data pada *cluster*  $C$ , dan  $d(i, j)$  adalah jarak antara data  $i$  dan  $j$ .

Setelah menghitung  $d(i, C)$  untuk semua  $C$ , maka diambil nilai terkecil menggunakan persamaan berikut :  $b(i) = \min_{C \neq A} d(i, C)$  (2.5)

$$b(i) = \min_{C \neq A} d(i, C) \quad (4)$$

Dimana  $b(i)$  adalah nilai terkecil dari jarak data ke semua objek pada *cluster* lainnya. 16

3. Hitung nilai *silhouette coefficient* menggunakan persamaan berikut :

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \quad (5)$$

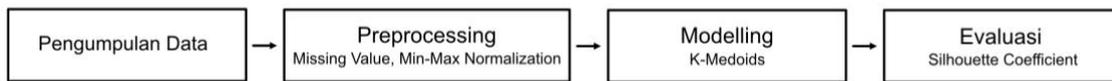
Nilai *silhouette coefficient* yang didapat berada dalam rentang -1 hingga 1 dimana semakin mendekati nilai 1, maka semakin bagus pengelompokkan objek-objek ke dalam sebuah *cluster* dan sebaliknya jika mendekati -1, maka akan makin buruk metode pengelompokkan datanya pada *cluster* tersebut.

Kriteria penilaian dan pengukuran baik tidaknya hasil *clustering* menurut Kaufman dan Roesseeuw dapat dilihat pada Tabel 1.

Nilai Silhouette Coefficient	Struktur
$0,7 < SC \leq 1$	Struktur Kuat
$0,5 < SC \leq 0,7$	Struktur Sedang
$0,25 < SC \leq 0,5$	Struktur Lemah
$SC \leq 0,25$	Tidak Terstruktur

**2. Research Method**

Metode yang digunakan pada penelitian ini adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM merupakan standar proses data mining yang ditujukan untuk melakukan proses analisis dari suatu industri sebagai dasar strategi pemecahan masalah dari bisnis atau unit penelitian [14]. Adapun tahapan umum yang dilakukan dimulai dari pengumpulan data yang akan digunakan. Kemudian, dilanjutkan dengan *preprocessing* untuk mempersiapkan data pada proses selanjutnya. Berikutnya dilakukan proses data mining untuk menemukan informasi penting dari data. Terakhir, melakukan evaluasi untuk mengetahui seberapa baik hasil dari proses data mining. Berikut adalah ilustrasi dari tahapan umum data mining.



Gambar 2. Tahapan Umum Data Mining

**3. Result and Analysis**

**3.1. Pengumpulan Data**

Data yang digunakan merupakan data sekunder yang didapatkan dari website resmi Badan Pusat Statistik Kabupaten Karawang ([karawangkab.bps.go.id](http://karawangkab.bps.go.id)) dan Google Maps. Adapun data yang diambil adalah kecamatan yang ada di karawang (30 kecamatan) dengan atribut faktor penyebab banjir berdasarkan penelitian sebelumnya, yaitu curah hujan, elevasi, kepadatan penduduk dan jarak ke sungai selama tahun 2019.

**3.2. Preprocessing**

*Preprocessing* merupakan proses untuk mempersiapkan data mentah menjadi data yang siap untuk dilakukan mining. Pada proses ini dilakukan dengan mengisi nilai yang hilang menggunakan nilai rata – rata. Selanjutnya, data dinormalisasi menggunakan *min-max normalization* untuk mengubah rentang data menjadi 1 – 0 dengan perhitungan menggunakan persamaan (1) sehingga setiap atribut sama pentingnya.

Tabel 2. Hasil Min-Max Normalization

Kecamatan	Elevasi	Kepadatan Penduduk	Curah Hujan	Jarak ke Sungai
Pangkalan	0,354	0,004	0,422	0,307
Tegalwaru	1,000	0,093	1,000	0,222
Ciampel	0,218	0,000	0,268	0,102
Telukjambe Timur	0,204	0,663	0,265	0,439
Telukjambe Barat	0,204	0,080	0,265	0,160
Klari	0,267	0,585	0,557	0,218
Cikampek	0,337	0,453	0,557	0,489
Purwasari	0,212	0,475	0,557	0,351
Tirtamulya	0,218	0,231	0,557	0,288
...	...	...	...	...
Cilebar	0,000	0,065	0,131	0,225
Cibuaya	0,008	0,045	0,139	0,165
Tirtajaya	0,043	0,079	0,276	0,000
Batujaya	0,023	0,100	0,456	0,047
Pakisjaya	0,043	0,074	0,275	0,010

**3.3. Data Mining**

Data yang telah dikumpulkan dan disiapkan kemudian akan dilakukan pengelompokkan menggunakan algoritma *k-medoids*. Algoritma ini dapat mengelompokkan data berdasarkan kemiripan satu sama lain. Berikut adalah proses yang dilakukan untuk mengelompokkan daerah rawan banjir di karawang.

1. Iterasi Pertama

a. Menentukan Jumlah *Cluster*

*Cluster* yang digunakan berjumlah 3 *cluster* yang terdiri dari *cluster* 1 (C1) merupakan daerah rawan banjir dengan potensi rendah, *cluster* 2 (C2) merupakan daerah rawan banjir dengan potensi sedang, dan *cluster* 3 (C3) merupakan daerah rawan banjir dengan potensi tinggi. Jumlah tersebut dipilih berdasarkan Indeks Risiko Bencana dari Badan Nasional Penanggulangan Bencana yang dibagi menjadi 3, yaitu rendah, sedang dan tinggi. Semakin tinggi potensi banjir, maka semakin tinggi juga risiko yang dirasakan.

b. Menentukan Pusat *Cluster*

Pusat *cluster* dipilih dengan menghitung bobot dari setiap atribut yang sudah dinormalisasi. Semakin tinggi kepadatan penduduk dan curah hujan, serta semakin rendah elevasi dan jarak ke sungai maka semakin tinggi potensi banjir di kecamatan tersebut. Sehingga, data yang dipilih menjadi pusat *cluster* 1 adalah kecamatan Kotabaru yang merupakan kecamatan dengan potensi banjir paling rendah, *cluster* 2 adalah kecamatan Cilamaya Kulon yang merupakan potensi banjir sedang, dan *cluster* 3 adalah kecamatan Karawang Barat yang merupakan potensi banjir paling tinggi.

Tabel 3. Pusat *Cluster* Iterasi Pertama

Cluster	Elevasi	Kepadatan Penduduk	Curah Hujan	Jarak ke Sungai
C1 (Rendah)	0,393	0,864	0,194	1,000
C2 (Sedang)	0,001	0,141	0,104	0,172
C3 (Tinggi)	0,204	0,990	0,201	0,197

c. Menghitung Jarak ke Pusat *Cluster*

Tahapan selanjutnya adalah menghitung jarak ke pusat *cluster*. Perhitungan jarak ini menggunakan persamaan *euclidean distance*. Berikut adalah perhitungan jarak data pertama menggunakan *euclidean distance*.

$$d(1,1) = \sqrt{(0,354 - 0,393)^2 + (0,004 - 0,864)^2 + (0,422 - 0,194)^2 + (0,307 - 1,000)^2} \\ = \sqrt{0,002 + 0,740 + 0,052 + 0,480} = \sqrt{1,274} = 1,129$$

$$d(1,2) = \sqrt{(0,354 - 0,001)^2 + (0,004 - 0,141)^2 + (0,422 - 0,104)^2 + (0,307 - 0,172)^2} \\ = \sqrt{0,125 + 0,019 + 0,101 + 0,018} = \sqrt{0,263} = 0,513$$

$$d(1,3) = \sqrt{(0,354 - 0,204)^2 + (0,004 - 0,990)^2 + (0,422 - 0,201)^2 + (0,307 - 0,197)^2} \\ = \sqrt{0,023 + 0,972 + 0,049 + 0,012} = \sqrt{1,056} = 1,028$$

Dari perhitungan tersebut didapatkan bahwa hasil jarak data pertama dengan pusat *cluster* pertama adalah 1,129, kemudian hasil jarak data pertama ke *cluster* kedua adalah 0,513, dan hasil jarak data pertama ke *cluster* ketiga adalah 1,028. Berdasarkan hasil ketiga perhitungan diatas, dapat diambil kesimpulan bahwa jarak data ke *cluster* 2 merupakan jarak yang paling dekat. Jarak lainnya dapat dilihat pada Tabel 4.

Tabel 4. Jarak Data ke Pusat *Cluster* Iterasi Pertama

Kecamatan	C1	C2	C3	Terdekat	Cluster
Pangkalan	1,129	0,513	1,028	0,513	2
Tegalwaru	1,489	1,344	1,442	1,344	2
Ciampel	1,260	0,314	0,997	0,314	2
Telukjambe Timur	0,629	0,642	0,412	0,412	3
Telukjambe Barat	1,166	0,266	0,913	0,266	2
Klari	0,915	0,689	0,544	0,544	3
Cikampek	0,751	0,719	0,720	0,719	2
Purwasari	0,858	0,627	0,645	0,627	2
Tirtamulya	1,034	0,523	0,844	0,523	2

...	...	...	...	...	...
Cilebar	1,182	0,097	0,950	0,097	2
Cibuaya	1,233	0,103	0,968	0,103	2
Tirtajaya	1,321	0,254	0,949	0,254	2
Batujaya	1,302	0,377	0,955	0,377	2
Pakisjaya	1,316	0,248	0,952	0,248	2
Total				9,114	

## 2. Iterasi Kedua

### a. Menentukan Pusat *Cluster*

Pemilihan pusat *cluster* pada iterasi kedua, *cluster* yang diubah hanya *cluster* 2 dikarenakan *cluster* 1 dan 3 merupakan kecamatan yang sudah pasti masuk *cluster* tersebut, begitu juga pada iterasi selanjutnya.

Tabel 5. Pusat *Cluster* Iterasi Kedua

Cluster	Elevasi	Kepadatan Penduduk	Curah Hujan	Jarak ke Sungai
C1 (Rendah)	0,393	0,864	0,194	1,000
C2 (Sedang)	0,038	0,182	0,114	0,246
C3 (Tinggi)	0,204	0,990	0,201	0,197

### b. Menghitung Jarak ke Pusat *Cluster*

Tabel 6. Jarak Data ke Pusat *Cluster* Iterasi Kedua

Kecamatan	C1	C2	C3	Terdekat	Cluster
Pangkalan	1,129	0,480	1,028	0,480	2
Tegalwaru	1,489	1,311	1,442	1,311	2
Ciampel	1,260	0,331	0,997	0,331	2
Telukjambe Timur	0,629	0,565	0,412	0,412	3
Telukjambe Barat	1,166	0,261	0,913	0,261	2
Klari	0,915	0,642	0,544	0,544	3
Cikampek	0,751	0,647	0,720	0,647	2
Purwasari	0,858	0,569	0,645	0,569	2
Tirtamulya	1,034	0,482	0,844	0,482	2
...	...	...	...	...	...
Cilebar	1,182	0,125	0,950	0,125	2
Cibuaya	1,233	0,164	0,968	0,164	2
Tirtajaya	1,321	0,312	0,949	0,312	2
Batujaya	1,302	0,404	0,955	0,404	2
Pakisjaya	1,316	0,305	0,952	0,305	2
Total				8,788	

### c. Menghitung Total Jarak Terdekat

$$\begin{aligned}
 S &= \text{total jarak terdekat baru} - \text{total jarak terdekat lama} \\
 &= 8,788 - 9,114 \\
 &= -0,326
 \end{aligned}$$

Karena nilai yang dihasilkan kurang dari 0, maka iterasi yang dilakukan masih terus berlanjut sampai menghasilkan nilai perbandingan total jarak terdekat baru dengan total jarak terdekat lama menghasilkan nilai lebih dari 0.

- 3. Iterasi Ketiga
  - a. Menentukan Pusat *Cluster*

Tabel 7. Pusat *Cluster* Iterasi Ketiga

Cluster	Elevasi	Kepadatan Penduduk	Curah Hujan	Jarak ke Sungai
C1 (Rendah)	0,393	0,864	0,194	1,000
C2 (Sedang)	0,053	0,196	0,136	0,280
C3 (Tinggi)	0,204	0,990	0,201	0,197

- b. Menghitung Jarak ke Pusat *Cluster*

Tabel 8. Jarak Data ke Pusat *Cluster* Iterasi Ketiga

Kecamatan	C1	C2	C3	Terdekat	Cluster
Pangkalan	1,129	0,459	1,028	0,459	2
Tegalwaru	1,489	1,288	1,442	1,288	2
Ciampel	1,260	0,339	0,997	0,339	2
Telukjambe Timur	0,629	0,532	0,412	0,412	3
Telukjambe Barat	1,166	0,260	0,913	0,260	2
Klari	0,915	0,615	0,544	0,544	3
Cikampek	0,751	0,607	0,720	0,607	2
Purwasari	0,858	0,535	0,645	0,535	2
Tirtamulya	1,034	0,454	0,844	0,454	2
...	...	...	...	...	...
Cilebar	1,182	0,151	0,950	0,151	2
Cibuaya	1,233	0,195	0,968	0,195	2
Tirtajaya	1,321	0,335	0,949	0,335	2
Batujaya	1,302	0,409	0,955	0,409	2
Pakisjaya	1,316	0,327	0,952	0,327	2
Total				8,793	

- c. Menghitung Total Jarak Terdekat

$$\begin{aligned}
 S &= \text{total jarak terdekat baru} - \text{total jarak terdekat lama} \\
 &= 8,793 - 8,788 \\
 &= 0,005
 \end{aligned}$$

Karena nilai yang dihasilkan lebih dari 0, maka iterasi yang dilakukan berhenti sampai disini dan hasil *cluster* yang dilakukan terdapat pada iterasi ketiga. Hasil *cluster* tersebut yaitu cluster 1 sebanyak 1 kecamatan, *cluster* 2 sebanyak 24 kecamatan dan *cluster* 3 sebanyak 5 kecamatan. Untuk masing-masing anggota *cluster* dapat dilihat pada tabel 9.

Tabel 9. Hasil Pengelompokkan

Cluster	Kecamatan	Potensi
1	Kotabaru.	Rendah
2	Pangkalan, Tegalwaru, Ciampel, Telukjambe barat, Cikampek, Purwasari, Tirtamulya, Jatisari, Banyusari, Cilamaya wetan, Cilamaya kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Tempuran, Kutawaluya, Jayakarta, Pedes, Cilebar, Cibuaya, Tirtajaya, Batujaya, Pakisjaya.	Sedang
3	Telukjambe timur, Klari, Karawang timur, Karawang barat, dan Rengasdengklok.	Tinggi

Setiap *cluster* dalam pada Tabel 9 mempunyai karakteristiknya masing-masing karena prinsip dari *clustering* adalah memaksimalkan kesamaan dalam satu *cluster* dan memaksimalkan perbedaan antar *cluster*. Karakteristik masing-masing *cluster* tersebut adalah sebagai berikut :

1. *Cluster* 1 mempunyai karakteristik dengan kepadatan penduduk tertinggi,
2. *Cluster* 2 mempunyai karakteristik dengan elevasi terendah dan jarak ke sungai terdekat,
3. *Cluster* 3 mempunyai karakteristik dengan curah hujan tertinggi.

### 3.4. Evaulasi

Hasil pengelompokkan dari algoritma *k-medoids* kemudian dilakukan evaluasi. Evaluasi ini dilakukan untuk mengetahui seberapa baik pengelompokkan yang telah dibuat. Evaluasi yang digunakan adalah *silhouette coefficient*. *Silhouette Coefficient* merupakan metode yang digunakan untuk melihat kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*.

Berikut ini langkah-langkah yang dilakukan untuk mendapatkan nilai *silhouette coefficient* :

1. Menghitung rata – rata jarak data terhadap data lain dalam satu *cluster*

Pengukuran ini dilakukan untuk mengetahui seberapa dekat satu data dengan data lainnya dalam satu *cluster*. Berikut ini perhitungan yang dilakukan untuk menghitung jarak data ke data lain dalam satu *cluster*.

$$\begin{aligned} a(\text{Kotabaru}) &= \frac{1}{1} \sqrt{(0,393 - 0)^2 + (0,864 - 0)^2 + (0,194 - 0)^2 + (1,000 - 0)^2} \\ &= \frac{1}{1} \sqrt{1,939} = 1,392 \end{aligned}$$

Pada *cluster* 1 anggota yang ada hanya satu data sehingga perhitungan jaraknya dikurangi 0 dan hasil dari  $a(\text{Kotabaru})$  sebesar 1,392. Hasil perhitungan lainnya dapat dilihat pada Tabel 10.

$$\begin{aligned} a(\text{Pangkalan}) &= \frac{1}{24 - 1} (\sqrt{(0,354 - 1,000)^2 + (0,004 - 0,093)^2 + (0,422 - 1,000)^2 + (0,307 - 0,222)^2} \\ &\quad + \sqrt{(0,354 - 0,218)^2 + (0,004 - 0,000)^2 + (0,422 - 0,268)^2 + (0,307 - 0,102)^2} \\ &\quad + \sqrt{(0,354 - 0,204)^2 + (0,004 - 0,080)^2 + (0,422 - 0,265)^2 + (0,307 - 0,160)^2} \\ &\quad + \dots \\ &\quad + \sqrt{(0,354 - 0,043)^2 + (0,004 - 0,074)^2 + (0,422 - 0,275)^2 + (0,307 - 0,010)^2}) \\ &= \frac{1}{23} (10,842) = 0,471 \end{aligned}$$

Pada *cluster* 2 anggota yang ada sebanyak 24 kecamatan sehingga perhitungan jarak diukur ke 23 kecamatan lainnya dalam satu *cluster* dan hasil dari  $a(\text{Pangkalan})$  sebesar 0,471. Hasil perhitungan lainnya dapat dilihat pada Tabel 10.

$$\begin{aligned} a(\text{Telukjambe Timur}) &= \frac{1}{5 - 1} (\sqrt{(0,204 - 0,267)^2 + (0,663 - 0,585)^2 + (0,265 - 0,557)^2 + (0,439 - 0,218)^2} \\ &\quad + \sqrt{(0,204 - 0,204)^2 + (0,663 - 1,000)^2 + (0,265 - 0,205)^2 + (0,439 - 0,236)^2} \\ &\quad + \sqrt{(0,204 - 0,204)^2 + (0,663 - 0,990)^2 + (0,265 - 0,201)^2 + (0,439 - 0,197)^2} \\ &\quad + \sqrt{(0,204 - 0,061)^2 + (0,663 - 0,689)^2 + (0,265 - 0,114)^2 + (0,439 - 0,188)^2}) \\ &= \frac{1}{4} (1,517) = 0,379 \end{aligned}$$

Pada *cluster* 3 anggota yang ada sebanyak 5 kecamatan sehingga perhitungan jarak diukur ke 4 kecamatan lainnya dalam satu *cluster* dan hasil dari  $a(\text{Telukjambe Timur})$  sebesar 0,379. Hasil lainnya pada perhitungan rata-rata jarak data terhadap data lain dalam satu *cluster* dapat dilihat pada Tabel 10.

2. Menghitung rata-rata jarak data dengan data lain yang berbeda *cluster*.

Pengukuran ini dilakukan untuk mengetahui seberapa jauh satu data dengan data lainnya yang berada di *cluster* berbeda.

$$\begin{aligned} d(\text{Kotabaru}, 2) &= \frac{1}{24} (\sqrt{(0,393 - 0,354)^2 + (0,864 - 0,004)^2 + (0,194 - 0,422)^2 + (1,000 - 0,307)^2} \\ &\quad + \sqrt{(0,393 - 1,000)^2 + (0,864 - 0,093)^2 + (0,194 - 1,000)^2 + (1,000 - 0,222)^2} \\ &\quad + \sqrt{(0,393 - 0,218)^2 + (0,864 - 0,000)^2 + (0,194 - 0,268)^2 + (1,000 - 0,102)^2} \\ &\quad + \dots \end{aligned}$$

$$\begin{aligned}
 & +\sqrt{(0,393 - 0,043)^2 + (0,864 - 0,074)^2 + (0,194 - 0,275)^2 + (1,000 - 0,010)^2} \\
 & = \frac{1}{24} (27,006) = 1,125
 \end{aligned}$$

$$\begin{aligned}
 d(\text{Kotabaru}, 3) &= \frac{1}{5} (\sqrt{(0,393 - 0,204)^2 + (0,864 - 0,663)^2 + (0,194 - 0,265)^2 + (1,000 - 0,439)^2} \\
 & +\sqrt{(0,393 - 0,267)^2 + (0,864 - 0,585)^2 + (0,194 - 0,557)^2 + (1,000 - 0,218)^2} \\
 & +\sqrt{(0,393 - 0,204)^2 + (0,864 - 1,000)^2 + (0,194 - 0,205)^2 + (1,000 - 0,236)^2} \\
 & +\sqrt{(0,393 - 0,204)^2 + (0,864 - 0,990)^2 + (0,194 - 0,201)^2 + (1,000 - 0,197)^2} \\
 & +\sqrt{(0,393 - 0,061)^2 + (0,864 - 0,689)^2 + (0,194 - 0,114)^2 + (1,000 - 0,188)^2}) \\
 & = \frac{1}{5} (4,075) = 0,815
 \end{aligned}$$

Dari perhitungan diatas, data Kotabaru yang diukur dengan objek pada *cluster* 2 menghasilkan nilai sebesar 1,125 dan data Kotabaru yang diukur dengan objek pada *cluster* 3 menghasilkan nilai sebesar 0,815. Setelah perhitungan  $d(i,C)$  diambil nilai terkecil untuk mendapatkan nilai  $b(i)$ . Hasil lainnya pada perhitungan rata-rata jarak data dengan data lain yang berbeda *cluster* dapat dilihat pada Tabel 10.

3. Menghitung nilai *silhouette coefficient*

Pengukuran nilai *silhouette coefficient* dilakukan untuk mengetahui kualitas dan kekuatan *cluster*, seberapa baik suatu objek ditempatkan dalam suatu *cluster*. Berikut ini perhitungan dari *silhouette coefficient*:

$$s(\text{kotabaru}) = \frac{0,815 - 1,392}{\max(0,815, 1,392)} = \frac{-0,577}{1,392} = -0,415$$

Dari perhitungan tersebut didapatkan nilai *silhouette coefficient* pada kecamatan Kotabaru sebesar - 0,415. Hasil perhitungan *silhouette coefficient* lainnya dapat dilihat pada Tabel 10.

Tabel 10. Hasil *Silhouette Coefficient*

Kecamatan	Cluster	a(i)	d(i, 1)	d(i, 2)	d(i, 3)	b(i)	s(i)
Kotabaru	1	1,392	-	1,125	0,815	0,815	-0,415
Pangkalan	2	0,471	1,129	-	0,839	0,839	0,438
Tegalwaru	2	1,190	1,489	-	1,308	1,308	0,090
Ciampel	2	0,377	1,260	-	0,829	0,829	0,545
Telukjambe Barat	2	0,332	1,166	-	0,743	0,743	0,553
Cikampek	2	0,628	0,751	-	0,555	0,555	-0,117
Purwasari	2	0,563	0,858	-	0,473	0,473	-0,159
Tirtamulya	2	0,463	1,034	-	0,653	0,653	0,290
Jatisari	2	0,411	1,091	-	0,619	0,619	0,335
...	...	...	...	...	...	...	...
Telukjambe Timur	3	0,379	0,629	0,628	-	0,628	0,396
Klari	3	0,493	0,915	0,607	-	0,607	0,188
Karawang Timur	3	0,470	0,799	0,892	-	0,799	0,412
Karawang Barat	3	0,335	0,834	0,883	-	0,834	0,599
Rengasdengklok	3	0,382	0,898	0,624	-	0,624	0,388
Rata - rata						0,370	

Dapat dilihat pada Tabel 10, pada kecamatan Kotabaru, Cikampek dan Purwasari nilai  $s(i)$  menghasilkan nilai negatif atau kurang dari 0. Hal tersebut dikarenakan nilai  $a(i)$  lebih besar dari nilai  $b(i)$  sehingga jika nilai  $s(i) < 0$  atau negatif, berarti data tersebut tidak tepat di *cluster* tersebut dan lebih tepat masuk *cluster* lainnya.

Secara keseluruhan nilai *silhouette coefficient* pada pengelompokan daerah rawan banjir di kabupaten karawang menggunakan *k-medoids* menghasilkan nilai sebesar 0,370. Hasil tersebut berdasarkan kriteria pengelompokan masuk ke dalam kategori *cluster* dengan *low structure* atau struktur lemah yang artinya tingkat kesamaan dalam satu *cluster* rendah dan tingkat kesamaan antar *cluster* tidak begitu jauh. Hal ini dapat terjadi karena perbedaan data yang sangat besar antara data satu dengan data lainnya dan juga bisa disebabkan karena belum ditemukan kombinasi *medoid* awal yang terbaik.

#### 4. Conclusion

Berdasarkan pengelompokan yang telah dilakukan menggunakan algoritma *k-medoids*, didapatkan kesimpulan sebagai berikut:

1. Daerah rawan banjir di Kabupaten Karawang dengan potensi banjir rendah sebanyak 1 kecamatan, yaitu Kotabaru. Potensi banjir sedang sebanyak 24 kecamatan yaitu Pangkalan, Tegalwaru, Ciampel, Telukjambe Barat, Cikampek, Purwasari, Tirtamulya, Jatisari, Banyusari, Cilamaya Wetan, Cilamaya Kulon, Lemahabang, Telagasari, Majalaya, Rawamerta, Tempuran, Kutawaluya, Jayakarta, Pedes, Cilebar, Cibuaya, Tirtajaya, Batujaya, Pakisjaya. Potensi banjir tinggi sebanyak 5 kecamatan yaitu Telukjambe Timur, Klari, Karawang Timur, Karawang Barat, Rengasdengklok.
2. Berdasarkan pengelompokan didapatkan faktor penyebab banjir di Karawang yang pertama adalah curah hujan, kemudian elevasi dan jarak ke sungai, dan terakhir kepadatan penduduk.
3. Penerapan algoritma *k-medoids* dalam menentukan daerah rawan banjir di Kabupaten Karawang mendapatkan nilai *silhouette coefficient* sebesar 0,370 yang masuk kategori *low structure*, namun masih dapat dijadikan sebagai acuan untuk melakukan upaya penanggulangan banjir lainnya.

#### References

- [1] Badan Meteorologi Klimatologi dan Geofisika, "Prakiraan Musim Hujan Tahun 2020/2021 di Indonesia". Retrieved 18 Juni 2021, from: <https://www.bmkg.go.id/berita/?p=prakiraan-musim-hujan-tahun-2020-2021-di-indonesia&lang=ID&s=detil>.
- [2] Dinas Komunikasi dan Informatika Kabupaten Karawang, "Statistik Sektorial Kabupaten Karawang 2020", Retrieved 18 Juni 2021, from: <https://karawangkab.go.id/dokumen/statistik-sektorial-kabupaten-karawang-2020>.
- [3] Riyanto, B., "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)", *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp 562-568. Oktober 2019.
- [4] Pramesti, D. F., Furqon, M. T., & Dewi, C., "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, pp 723-732, Juni 2017.
- [5] Simamora, D. A. S., Furqon, M. T., & Priyambadha, B., "Clustering Data Kejadian Tsunami Yang Disebabkan Oleh Gempa Bumi Dengan Menggunakan Algoritma K-Medoids", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 8, pp 635-640. Juni 2017.
- [6] Khusaeri, A., Ilham, S., et al., "Algoritma C4.5 Untuk Pemodelan Daerah Rawan Banjir Studi Kasus Kabupaten Karawang Jawa Barat", *ILKOM Jurnal Ilmiah*, vol. 9, no. 2, pp 132-136, Agustus 2017.
- [7] Lutfi, M., & Hasyim, M., "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-NN Imputation Pada Algoritma C4.5", *Jurnal Rekayasa Sistem Komputer*, vol. 2, no. 2. Oktober 2019.
- [8] Nasution, D. A., Khotimah, H. H., Chamidah. N., "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN", *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, pp. 78-82. Januari 2019.
- [9] Riyanto, B., "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)", *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp 562-568. Oktober 2019.
- [10] Anggreini, N. L.; Tresnawati, S.; , "Komparasi Algoritma K-Means dan K-Medoids untuk Menangani Strategi Promosi di Politeknik TEDC Bandung", *Jurnal TEDC*, vol. 4, no. 2, pp 120-127. Mei 2020.
- [11] Han, J., Kamber, M., & Pei, J., "*Data Mining Concepts and Techniques Third Edition*", Waltham: Elsevier. 2012.
- [12] Anggara, M., Sujiani, H., & Nasution, H., "Pemilihan Distance Measure pada K-Means Clustering untuk Pengelompokan Member di Alvaro Fitness", *Jurnal Sistem dan Teknologi Informasi*, vol. 1, no. 1, pp. 1-6. 2016.
- [13] Sari, K., "*Analisis Evaluasi Perhitungan Jarak Terhadap Nilai Silhouette Coefficient pada Algoritma K-Means*", [Tesis], Medan: Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara. 2020.
- [14] Adiana, B. E., Soesanti, I., & Permanasari, A. E., "Analisis Segmentasi Pelanggan Menggunakan Kombinasi RFM Model dan Teknik Clustering", *Jurnal Terapan Teknologi Informasi*, vol. 2, no. 1, pp 23-32. April 2018.