

IMPLEMENTASI ALGORITME *K-MEANS CLUSTERING* UNTUK PENGELOMPOKKAN WILAYAH RAWAN BANJIR

Try Wathoriq^{1*}, Subandi²

^{1,2}Fakultas Teknologi Informasi, Teknik Informatika, Universitas Budi Luhur, Jakarta, Indonesia

Email: ^{1*}trwathoriq12@gmail.com, ²subandi@budiluhur.ac.id

(* : corresponding author)

Abstrak- Badan Penanggulangan Bencana Daerah (BPBD) Kota Tangerang menghadapi tantangan dalam mengelola data bencana banjir yang sering terjadi di wilayah tersebut. Dalam upaya untuk meningkatkan efisiensi dan efektivitas penanganan bencana banjir, diperlukan pendekatan yang efisien untuk mengelompokkan wilayah berdasarkan tingkat dampak yang dialami. Data mining, dengan metode K-Means Clustering, dapat menjadi solusi yang potensial untuk mengidentifikasi pola dan pengelompokan wilayah banjir yang serupa berdasarkan dampak yang terjadi. Penelitian ini bertujuan untuk mengimplementasikan metode data mining K-Means Clustering dalam pengelompokkan wilayah banjir berdasarkan dampak yang terjadi pada data yang diberikan BPBD Kota Tangerang. Penelitian ini diharapkan dapat menghasilkan pengelompokan wilayah banjir yang lebih efisien dalam upaya pengambilan keputusan dalam penanganan bencana banjir. Data historis tentang bencana banjir di Kota Tangerang dikumpulkan dan dipersiapkan untuk analisis. Metode K-Means Clustering diterapkan untuk mengelompokkan wilayah berdasarkan tingkat dampak yang dialami akibat banjir. Pengelompokan ini dilakukan dengan mempertimbangkan parameter seperti jumlah kejadian, jumlah kerusakan, korban menderita, korban meninggal, korban menghilang, korban mengungsi. Evaluasi dilakukan untuk mengukur kualitas pengelompokan dan mendapatkan wawasan yang berarti. Pada hasil penelitian ini Implementasi data mining K-Means Clustering berhasil menghasilkan pengelompokan wilayah banjir yang signifikan berdasarkan tingkat dampak yang terjadi. Wilayah dengan dampak banjir serupa dikelompokkan bersama, memungkinkan BPBD Kota Tangerang untuk mengidentifikasi pola dan karakteristik yang berkaitan dengan tingkat dampak. Tujuan utama dari penelitian ini adalah untuk menganalisis dampak yang terjadi akibat banjir pada wilayah yang rentan terhadap bencana banjir, melalui implementasi teknik data mining menggunakan algoritme K-Means Clustering. Penelitian ini bertujuan untuk mengidentifikasi potensi kerawanan banjir berdasarkan dampak yang terjadi. Dengan adanya penelitian ini, peneliti berharap dapat memberikan wawasan yang lebih mendalam tentang dampak banjir pada wilayah yang rawan, serta membantu dalam mengembangkan langkah-langkah konkrit untuk mengurangi kerugian dan merencanakan tindakan yang lebih adaptif dalam menghadapi risiko banjir.

Kata Kunci: Banjir, Data Mining, Clustering, K-Means.

IMPLEMENTATION OF *K-MEANS CLUSTERING* ALGORITHM FOR *CLUSTERING FLOOD-PRONE AREA*

Abstract- The Tangerang City Regional Disaster Management Agency (BPBD) faces challenges in managing data on flood disasters that frequently occur in the region. In an effort to improve the efficiency and effectiveness of flood disaster management, an efficient approach is needed to cluster areas based on the level of impact experienced. Data mining, using the K-Means Clustering method, can be a potential solution to identify patterns and group similar flood areas based on the impact. This research aims to implement the K-Means Clustering data mining method in clustering flood areas based on the impact on the data provided by BPBD Tangerang City. This research is expected to produce a more efficient grouping of flood areas in an effort to make decisions in handling flood disasters. Historical data on flood disasters in Tangerang City was collected and prepared for analysis. The K-Means Clustering method was applied to group the areas based on the level of impact experienced due to flooding. This clustering was done by considering parameters such as number of events, number of damages, victims suffered, victims died, victims disappeared, victims displaced. Evaluation is done to measure the quality of the clustering and gain meaningful insights. In the results of this study, the implementation of K-Means Clustering data mining successfully resulted in significant clustering of flood areas based on the level of impact. Areas with similar flood impacts were grouped together, allowing BPBD Tangerang City to identify patterns and characteristics related to the level of impact. The main objective of this research is to analyse the impact of flooding on flood-prone areas, through the implementation of data mining techniques using the K-Means Clustering algorithm. This research aims to identify potential flood vulnerability based on the impacts that occur. With this study, the researcher hopes to provide a deeper insight into the impact of flooding on vulnerable areas, as well as assist in developing concrete steps to reduce losses and plan more adaptive actions in the face of flood risk.

Keywords: Flooding, Data Mining, Clustering, K-Means.

1. PENDAHULUAN

Menghadapi bencana banjir yang sering terjadi di wilayah tersebut menyulitkan Badan Penanggulangan Bencana Daerah (BPBD) Kota Tangerang. Metode klasifikasi wilayah yang efisien berdasarkan tingkat keparahan dampak diperlukan untuk meningkatkan efektivitas dan efisiensi penanggulangan bencana banjir. Menemukan pola dan mengklasifikasikan lokasi banjir yang dapat diterima berdasarkan dampak yang dihadapi dapat dicapai dengan menggunakan data mining dan pendekatan pengelompokan K-Means. Teknologi penambangan data telah maju dalam beberapa tahun terakhir menjadi alat yang ampuh untuk menggunakan volume data yang sangat besar untuk mengungkap pola dan tren tersembunyi. K-means clustering merupakan salah satu teknik yang sering digunakan dalam data mining. Tujuan teknik ini adalah untuk mengklasifikasikan item data ke dalam kategori yang dapat dibandingkan berdasarkan kesamaan sifat. Menerapkan k-means clustering data mining untuk analisis banjir dapat memberikan manfaat yang signifikan. Dengan mengumpulkan dan menganalisis data historis banjir.

Pada penelitian sebelumnya yang berjudul "KMeans Clustering Menggunakan RapidMiner dalam Segmentasi Pelanggan dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal" [1], terdapat perbedaan yang didapatkan setelah melakukan peninjauan lebih lanjut, bahwa pada penelitian tersebut proses yang dilakukan untuk melihat segmentasi pelanggan menggunakan sebuah tools (Rapid Miner), tidak dalam sebuah sistem yang di rancang sedemikian rupa. Sehingga hasil output dari perhitungan yang sudah dilakukan tidak dapat di konversikan menjadi sebuah dokumen.

Penambangan data, menurut studi sebelumnya juga, adalah proses menggunakan pendekatan Pohon Keputusan Teknis untuk menemukan pola atau informasi yang menarik dalam data yang dipilih. Pendekatan atau algoritme terbaik untuk digunakan benar-benar bergantung pada tujuan menyeluruh proses KDD [2]. Menemukan pola spesifik dalam kumpulan data besar adalah proses yang disebut penambangan data. Penambangan data adalah teknik yang dapat digunakan dan mempermudah pekerjaan di domain lain juga, meskipun telah dieksplorasi secara luas di bidang ilmu komputer dan statistik [3]. K-means adalah metode yang mengalokasikan nilai cluster (k) secara acak. Untuk sementara, nilainya berperan sebagai centroid, atau pusat cluster. Kemudian, dengan menggunakan metode tersebut, tentukan jarak antara setiap potongan data yang ada dan setiap centroid hingga nilai centroid tetap konstan [4]. Dengan berulang kali menghitung jarak antara setiap titik dan pusat cluster, pendekatan clustering K-Means dihasilkan dari sebuah dataset [3].

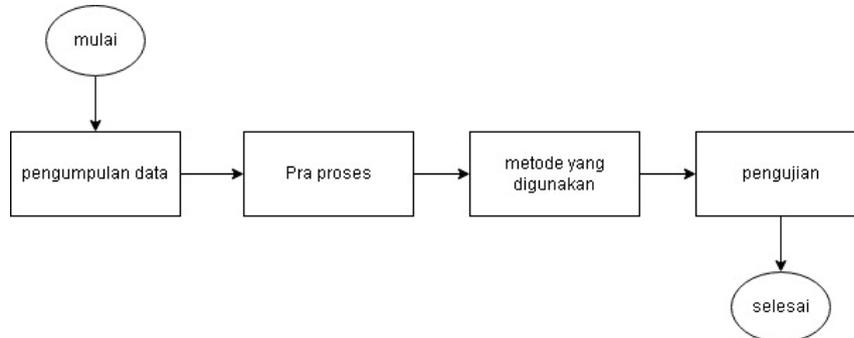
Suatu teknik pengelompokan berdasarkan kedekatan (similarity) disebut clustering. Jika istilah "kelompok" mengacu pada sekelompok orang yang berbagi keadaan yang sama, pengelompokan berbeda dari kelompok. Pengelompokan tersebut didasarkan pada kesamaan karakteristik sampel yang ada, salah satunya dengan memanfaatkan rumus jarak Euclidean [5]. Pengelompokan kelompok tidak harus sama. Teknik text mining seperti clustering digunakan untuk membagi objek atau data menjadi beberapa kelompok (cluster) sedemikian rupa sehingga data yang terkait akan berada dalam satu cluster dan jarak antar cluster akan dibuat sejauh mungkin. Karena kesederhanaan dan tingkat akurasinya yang tinggi sehubungan dengan ukuran data, K Means Clustering adalah algoritma yang populer untuk masalah pengelompokan. Selain itu, algoritma ini tidak terpengaruh oleh urutan objek sehingga lebih efektif dan terukur saat memproses data dalam jumlah besar [6]. Mencari tahu berapa banyak cluster data yang layak dipisahkan selama validasi adalah tujuannya. Indeks Davies-Bouldin adalah salah satu dari banyak metode untuk validasi klaster. Cluster optimal dalam dataset yang diuji ditentukan melalui validasi clustering [7]. DBI adalah salah satu teknik penilaian klaster pada pendekatan pengelompokan berbasis nilai kohesi dan pemisahan. Kohesi dalam sebuah clustering adalah total jarak data dari centroid cluster yang sedang dilacak. Untuk selanjutnya pemisahan ditentukan oleh jarak antar centroid cluster [8]. Metode yang dikenal dengan Knowledge Discovery in Databases (KDD) mencoba menelaah dan mengevaluasi beberapa set data untuk mendapatkan pengetahuan dan informasi yang relevan [9].

Tujuan dari penelitian ini adalah untuk merancang dan membangun sebuah sistem untuk menganalisis data bencana untuk mengidentifikasi daerah yang rawan banjir dalam data yang disediakan oleh BPBD Kota Tangerang dengan menggunakan metode clustering dan algoritma K-means untuk meningkatkan kesadaran dan implementasi yang efektif. mitigasi bencana.

2. METODE PENELITIAN

Teknik penelitian berfungsi sebagai kerangka acuan untuk melakukan penelitian, menjadikannya bagian penting dari proses. Tahap peneliti selanjutnya, yang berdasarkan **Gambar 1**, dimulai dengan mengumpulkan informasi dari subjek. Preprocessing adalah fase berikut, yang melibatkan memodifikasi data sesuai dengan

informasi yang diperlukan. Prosedur untuk menemukan nilai cluster dan data grup muncul berikutnya. Pengujian adalah fase terakhir.



Gambar 1. Metode Penelitian.

2.1 Pengumpulan Data

Tahap penelitian yang peneliti lakukan sambil mengumpulkan data dikenal sebagai tahap pengumpulan data. Pada tahap penelitian ini dilakukan observasi lapangan untuk melihat keadaan, permasalahan, dan prosedur pengumpulan data di lokasi penelitian BPBD Kota Tangerang. Data yang diterima adalah data history bencana dari BPBD Kota Tangerang. Data bencana tersebut merupakan data yang diambil dalam rentang waktu 1 tahun terakhir semenjak July 2023 kebelakang. Data yang digunakan terdiri dari 13 kecamatan dan data observasi dampak bencana banjir.

2.1.1 Data kebencanaan

Tabel 1. Data Dampak Bencana Banjir.

Kecamatan	Banyaknya Kejadian	Banyaknya Kerusakan	Korban Menderita	Korban Meninggal	Korban Hilang	Mengungsi
Batu Ceper	3	258	329	1	0	670
Benda	1	19	203	2	0	48
Cibodas	2	15	225	2	2	38
Ciledug	2	321	589	1	1	834
Cipondoh	1	136	150	1	0	354
Jatiuwung	2	155	319	0	1	402
Karang Tengah	8	479	601	3	2	1245
Karawaci	3	217	200	1	0	564
Larangan	4	369	502	3	2	960
Neglasari	2	152	564	2	1	396
Periuk	2	53	367	1	3	139
Pinang	1	264	325	2	1	687
Tangerang	2	93	346	0	0	242

Pada Tabel 1 dijelaskan didalamnya berisi data bencana banjir yang telah di olah kembali oleh BPBD Kota Tangerang. Data tersebut merupakan riwayat kejadian banjir dalam rentang waktu 1 tahun terakhir sejak juli 2023 kebelakang.

2.2 Pra Pemrosesan.

Pra pemrosesan adalah tahapan yang dilakukan peneliti sebelum memproses data yang digunakan sebagai bahan penelitian. Tahapan ini bertujuan untuk menghilangkan beberapa permasalahan yang bisa mengganggu saat pemrosesan data dan memastikan kualitas data baik sebelum digunakan saat analisis data. Dalam proses ini, peneliti sudah mendapatkan data yang telah di olah sebelumnya oleh BPBD Kota Tangerang untuk di proses pada sistem yang dibuat ini.

2.3 Metode Yang Digunakan.

2.3.1 Clustering

Clustering adalah metode pengelompokan berdasarkan ukuran kedekatan (similarity). *Clustering* berbeda dengan group, jika group berarti group yang memiliki kondisi yang sama. Sedangkan pengelompokan cluster tidak harus sama, pengelompokan didasarkan pada kedekatan karakteristik sampel yang ada, salah satunya dengan menggunakan rumus jarak Euclidean [5]. *Clustering* adalah salah satu metode data mining yang digunakan untuk mengelompokkan objek atau data menjadi beberapa kelompok (cluster sehingga data yang serupa akan berada dalam satu cluster dan membuat jarak antar cluster sejauh mungkin [10]. K-Means merupakan suatu algoritma pengklasteran yang cukup sederhana yang mempartisi database ke dalam beberapa cluster k. Algoritma cukup mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah disesuaikan dan banyak digunakan. Prinsip utama dari teknik ini adalah menyusun K buah partisi/pusat massa (centroid)/rata-rata (mean) dari sekumpulan data. K Means *Clustering* adalah algoritma yang biasa digunakan dalam *clustering* masalah karena kesederhanaannya dan memiliki tingkat akurasi yang tinggi terhadap ukuran data sehingga K Means *Clustering* lebih efisien dan terukur dalam memproses data dalam jumlah besar, selain itu urutan objek tidak mempengaruhi algoritma ini [6] melakukan proses klaster dengan mengambil jarak terpendek dari setiap data yang diproses menggunakan rumus sebagai berikut:

$$D(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2}$$

Keterangan :

- D(x,y) adalah Euclidean Distance
- (x) merupakan koordinat object
- (y) merupakan koordinat centroid

2.3.2 Algoritme K-Means.

K-means adalah metode yang mengalokasikan nilai cluster (k) secara acak. Untuk sementara, nilainya berperan sebagai centroid, atau pusat cluster. Kemudian, dengan menggunakan metode tersebut, tentukan jarak antara setiap potongan data yang ada dan setiap centroid hingga nilai centroid tetap konstan [16]. Dengan berulang kali menghitung jarak antara setiap titik dan pusat cluster, pendekatan clustering K-Means dihasilkan dari sebuah dataset [2]. Algoritma K-Means berisi aturan berikut:

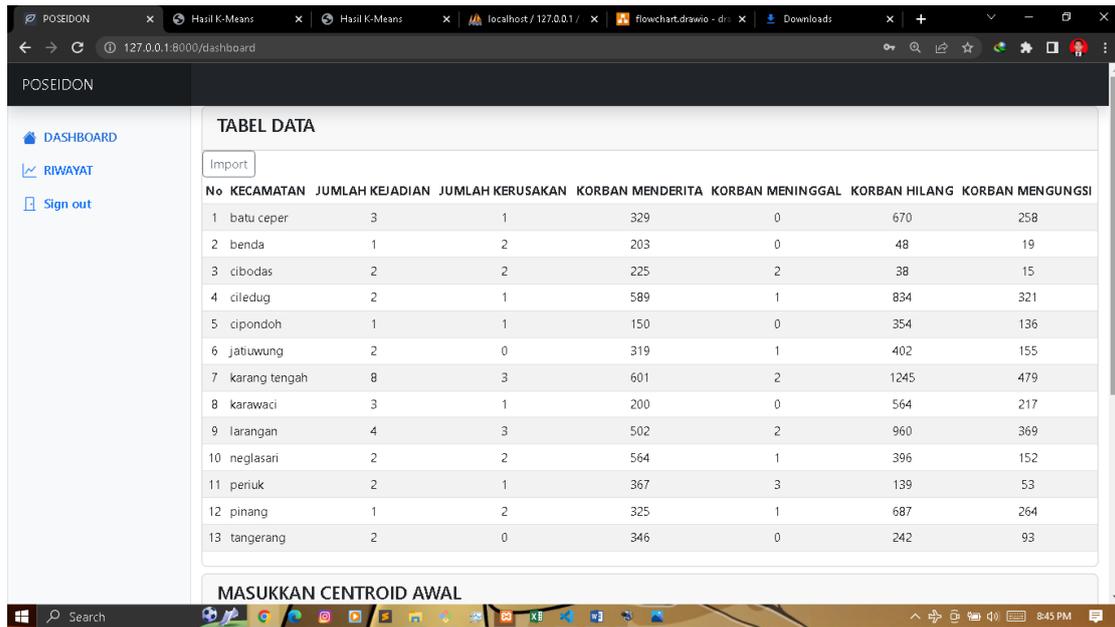
- 1) Total cluster yang dibutuhkan.
- 2) Jenis atribut adalah numeric. Proses untuk *clustering* menggunakan algoritme k-means adalah sebagai berikut :
 1. Tetapkan nilai k untuk keseluruhan cluster yang ingin Anda buat.
 2. Dimulai dengan k, pilih nilai cluster centroid (centroid).
 3. Untuk menentukan jarak terpendek antara setiap titik data dan setiap centroid, hitung jarak antara setiap titik data masukan dan setiap centroid menggunakan algoritma jarak Euclidean.
 4. Urutkan atau kelompokkan setiap potongan data sesuai dengan jaraknya dari centroid.
 5. Perbarui mediannya. nilai inti segar.
 6. Tindakan berulang Setiap 3–5 anggota cluster tidak berubah.

2.4 Pengujian.

Tahapan pengujian dilakukan untuk menemukan kelompok dari masing-masing kecamatan di kota Tangerang dari mulai yang paling berpotensi terkena banjir hingga yang rendah potensinya. Penerapan algoritma K-Means dilakukan menggunakan Euclidean Distance dengan menentukan centroid (titik pusat cluster) untuk melakukan perhitungan iterasi. Yang nantinya dilanjut dengan menghitung rasio dan melakukan perhitungan evaluasi menggunakan Davies Bouldin Index sebagai pengukur ke akuratan data yang di proses.

3. HASIL DAN PEMBAHASAN

Gambar 2 adalah hasil import dataset yang berasal dari data yang di berikan oleh BPBD Kota Tangerang yang nantinya akan di proses dengan cara memilih centroid awal pada halaman utama lalu dilanjut dengan memprosesnya.



Gambar 2 Input dataset

Tabel 2 Centroid Awal

Kecamatan	Banyaknya kejadian	Korban Meninggal	Korban Menderita	Korban Hilang	Mengungsi	Banyaknya Kerusakan
Batu Ceper	3	1	329	0	670	258
Karang Tengah	8	3	601	2	1245	479
Tangerang	2	0	346	0	242	258

Proses perhitungan awal menggunakan Euclidean Distance dimulai dengan mencari pusat cluster pertama secara random seperti yang sudah dijelaskan pada proses input centroid awal sebelumnya. seperti yang di paparkan pada **Tabel 2** Dan user memilih batu ceper, karang tengah, dan Tangerang. Dengan melakukan perhitungan jarak setiap data ke centroid, akan terbentuk perhitungan dengan rumus sebagai berikut.

$$(d) = \sqrt{(x2 - x1)^2 + (y2 - y1)^2 + \dots + (n2 - n1)^2}$$

mulai menghitung iterasi 1.

1. data 1 (Batu Ceper)

a) jarak data 1 → cluster 1

$$d = \sqrt{(3 - 3)^2 + (1 - 1)^2 + (329 - 329)^2 + (0 - 0)^2 + (670 - 670)^2 + (258 - 258)^2}$$

$$d = 0$$

b) jarak data 1 → cluster 2

$$d = \sqrt{(3 - 8)^2 + (1 - 3)^2 + (329 - 601)^2 + (0 - 2)^2 + (670 - 1245)^2 + (258 - 479)^2}$$

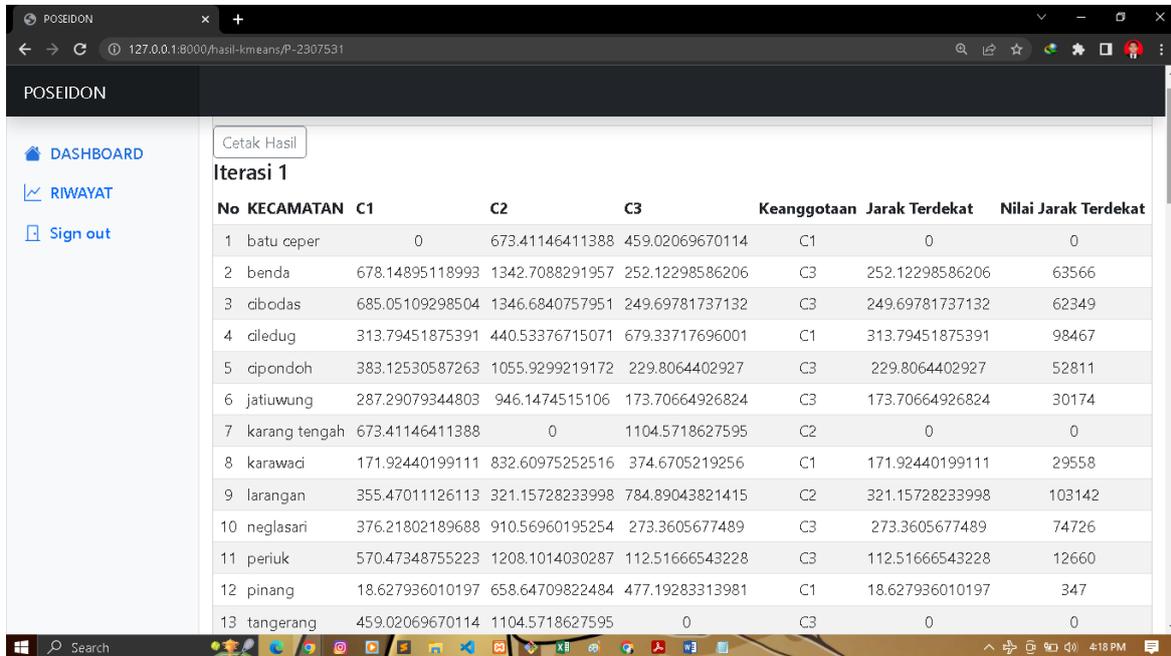
$$d = 673,41146$$

c) jarak data 1 → cluster 3

$$d = \sqrt{(3 - 2)^2 + (1 - 0)^2 + (329 - 346)^2 + (0 - 0)^2 + (670 - 242)^2 + (258 - 93)^2}$$

$$d = 459,0207$$

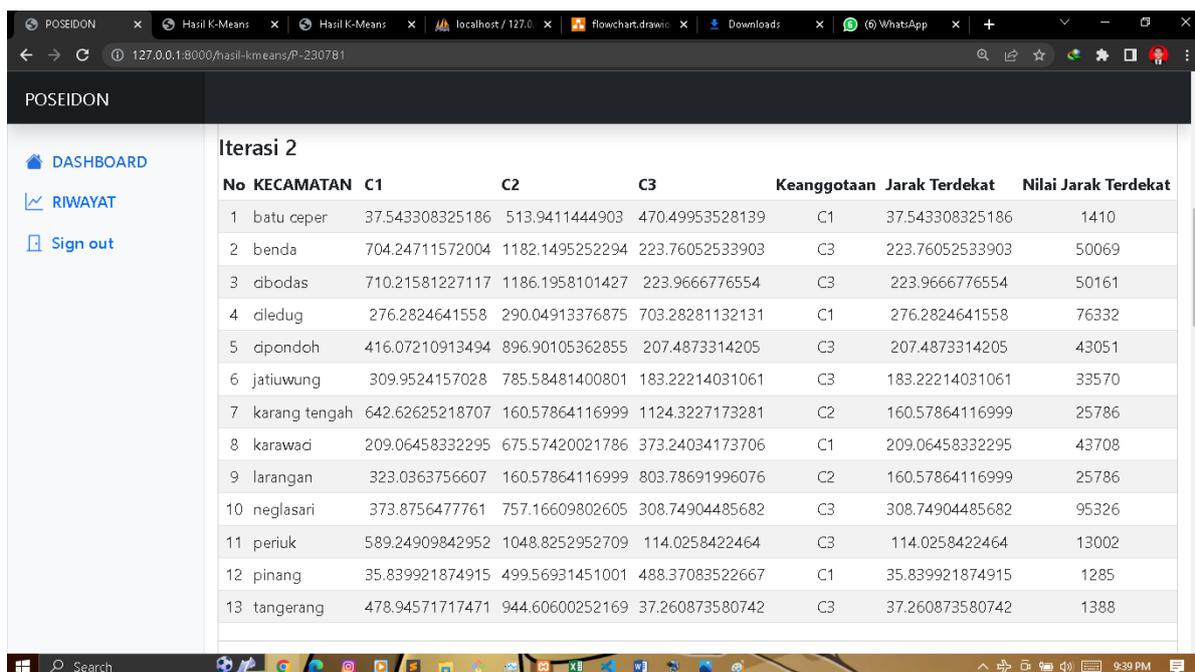
Dan perhitungan dilanjut hingga data ke-13 dengan pola dan cara perhitungan seperti diatas, apabila perhitungan telah dilakukan hingga data ke-13, maka dilanjutkan untuk mencari keanggotaan cluster dengan cara mencari jarak terdekat (jumlah terkecil). Pada tahap ini peneliti menggunakan Euclidean Distance untuk melakukan iterasi di atas. Dan berikut adalah hasil perhitungannya pada aplikasi.



No	KECAMATAN	C1	C2	C3	Keanggotaan	Jarak Terdekat	Nilai Jarak Terdekat
1	batu ceper	0	673.41146411388	459.02069670114	C1	0	0
2	benda	678.14895118993	1342.7088291957	252.12298586206	C3	252.12298586206	63566
3	cibodas	685.05109298504	1346.6840757951	249.69781737132	C3	249.69781737132	62349
4	ciledug	313.79451875391	440.53376715071	679.33717696001	C1	313.79451875391	98467
5	cipondoh	383.12530587263	1055.9299219172	229.8064402927	C3	229.8064402927	52811
6	jatiuwung	287.29079344803	946.1474515106	173.70664926824	C3	173.70664926824	30174
7	karang tengah	673.41146411388	0	1104.5718627595	C2	0	0
8	karawaci	171.92440199111	832.60975252516	374.6705219256	C1	171.92440199111	29558
9	larangan	355.47011126113	321.15728233998	784.89043821415	C2	321.15728233998	103142
10	neglasari	376.21802189688	910.56960195254	273.3605677489	C3	273.3605677489	74726
11	periuk	570.47348755223	1208.1014030287	112.51666543228	C3	112.51666543228	12660
12	pinang	18.627936010197	658.64709822484	477.19283313981	C1	18.627936010197	347
13	tangerang	459.02069670114	1104.5718627595	0	C3	0	0

Gambar 3 Iterasi 1

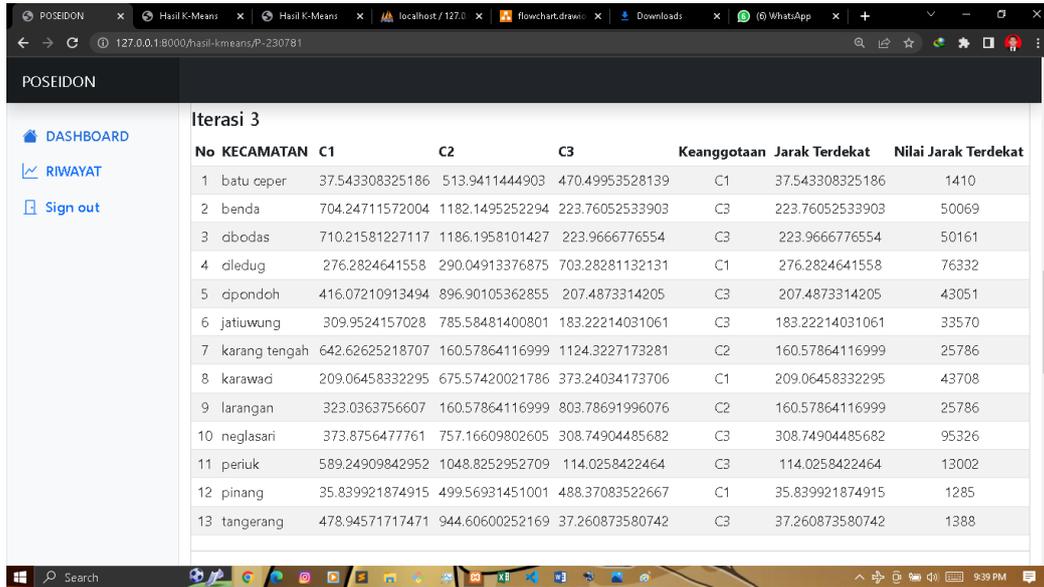
Terlihat jika di perhatikan pada Gambar 4, keanggotaannya masih berubah, maka harus dilakukan perhitungan iterasi ke 3.



No	KECAMATAN	C1	C2	C3	Keanggotaan	Jarak Terdekat	Nilai Jarak Terdekat
1	batu ceper	37.543308325186	513.9411444903	470.49953528139	C1	37.543308325186	1410
2	benda	704.24711572004	1182.1495252294	223.76052533903	C3	223.76052533903	50069
3	cibodas	710.21581227117	1186.1958101427	223.9666776554	C3	223.9666776554	50161
4	ciledug	276.2824641558	290.04913376875	703.28281132131	C1	276.2824641558	76332
5	cipondoh	416.07210913494	896.90105362855	207.4873314205	C3	207.4873314205	43051
6	jatiuwung	309.9524157028	785.58481400801	183.22214031061	C3	183.22214031061	33570
7	karang tengah	642.62625218707	160.57864116999	1124.3227173281	C2	160.57864116999	25786
8	karawaci	209.06458332295	675.57420021786	373.24034173706	C1	209.06458332295	43708
9	larangan	323.0363756607	160.57864116999	803.78691996076	C2	160.57864116999	25786
10	neglasari	373.8756477761	757.16609802605	308.74904485682	C3	308.74904485682	95326
11	periuk	589.24909842952	1048.8252952709	114.0258422464	C3	114.0258422464	13002
12	pinang	35.839921874915	499.56931451001	488.37083522667	C1	35.839921874915	1285
13	tangerang	478.94571717471	944.60600252169	37.260873580742	C3	37.260873580742	1388

Gambar 4 Iterasi 2

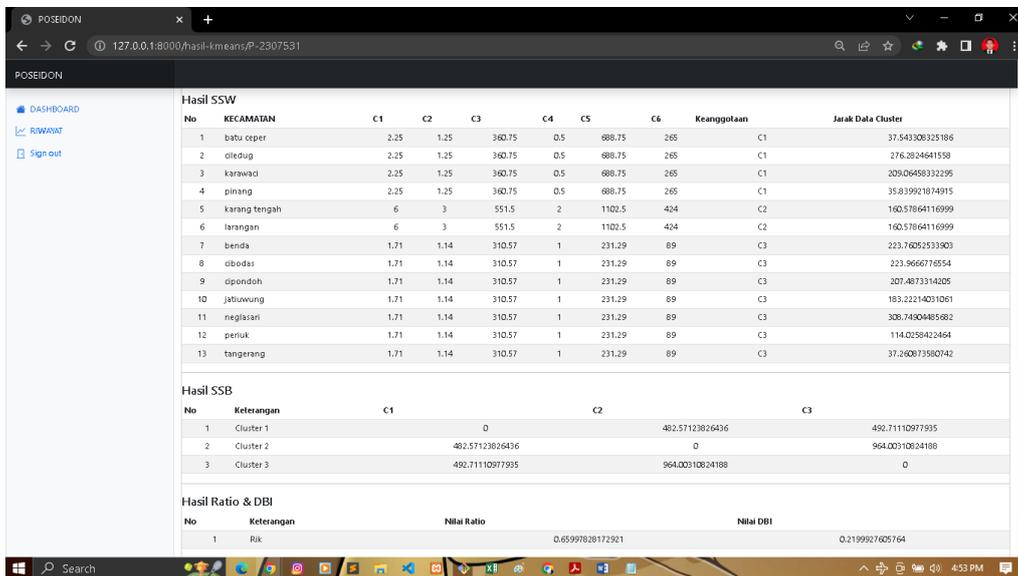
Pada Gambar 5, terlihat keanggotaan sudah tidak berubah dan sama seperti pada iterasi ke-2.



Iterasi 3							
No	KECAMATAN	C1	C2	C3	Keanggotaan	Jarak Terdekat	Nilai Jarak Terdekat
1	batu ceper	37.543308325186	513.9411444903	470.49953528139	C1	37.543308325186	1410
2	benda	704.24711572004	1182.1495252294	223.76052533903	C3	223.76052533903	50069
3	cibodas	710.21581227117	1186.1958101427	223.9666776554	C3	223.9666776554	50161
4	ciledug	276.2824641558	290.04913376875	703.28281132131	C1	276.2824641558	76332
5	dipondoh	416.07210913494	896.90105362855	207.4873314205	C3	207.4873314205	43051
6	jatiuwung	309.9524157028	785.58481400801	183.22214031061	C3	183.22214031061	33570
7	karang tengah	642.62625218707	160.57864116999	1124.3227173281	C2	160.57864116999	25786
8	karawad	209.06458332295	675.57420021786	373.24034173706	C1	209.06458332295	43708
9	larangan	323.0363756607	160.57864116999	803.78691996076	C2	160.57864116999	25786
10	neglasari	373.8756477761	757.16609802605	308.74904485682	C3	308.74904485682	95326
11	peruk	589.24909842952	1048.8252952709	114.0258422464	C3	114.0258422464	13002
12	pinang	35.839921874915	499.56931451001	488.37083522667	C1	35.839921874915	1285
13	tangerang	478.94571717471	944.60600252169	37.260873580742	C3	37.260873580742	1388

Gambar 5 Iterasi 3

Maka ketika nilai cluster di iterasi ke 3 sudah tak berubah, maka bisa melanjutkan ke perhitungan selanjutnya untuk sesuai KDD yakni sesi Evaluation menggunakan Davies Bouldin Index. Ada 4 tahap dalam perhitungan DBI yakni menghitung SSW, SSB, Rasio, DBI. Berikut gambaran hasil pada aplikasi.



Hasil SSW									
No	KECAMATAN	C1	C2	C3	C4	C5	C6	Keanggotaan	Jarak Data Cluster
1	batu ceper	2.25	1.25	360.75	0.5	888.75	265	C1	37.543308325186
2	ciledug	2.25	1.25	360.75	0.5	888.75	265	C1	276.2824641558
3	karawad	2.25	1.25	360.75	0.5	888.75	265	C1	209.06458332295
4	pinang	2.25	1.25	360.75	0.5	888.75	265	C1	35.839921874915
5	karang tengah	6	3	551.5	2	1102.5	424	C2	160.57864116999
6	larangan	6	3	551.5	2	1102.5	424	C2	160.57864116999
7	benda	1.71	1.14	310.57	1	231.29	89	C3	223.76052533903
8	cibodas	1.71	1.14	310.57	1	231.29	89	C3	223.9666776554
9	dipondoh	1.71	1.14	310.57	1	231.29	89	C3	207.4873314205
10	jatiuwung	1.71	1.14	310.57	1	231.29	89	C3	183.22214031061
11	neglasari	1.71	1.14	310.57	1	231.29	89	C3	308.74904485682
12	peruk	1.71	1.14	310.57	1	231.29	89	C3	114.0258422464
13	tangerang	1.71	1.14	310.57	1	231.29	89	C3	37.260873580742

Hasil SSB				
No	Keterangan	C1	C2	C3
1	Cluster 1	0	0	492.57123826436
2	Cluster 2	492.57123826436	0	964.00310824188
3	Cluster 3	492.57123826436	964.00310824188	0

Hasil Ratio & DBI			
No	Keterangan	Nilai Ratio	Nilai DBI
1	Rik	0.65997828172921	0.2199927605764

Gambar 6 Evaluasi Davies Bouldin Index.

Pada Gambar 6 sudah bisa di simpulkan bahwasannya tiap-tiap kecamatan sudah memiliki nilai cluster masing-masing berdasarkan keanggotaannya. Dan berikut adalah hasil pengelompokkannya.

Tabel 3 Hasil Pengelompokkan Kecamatan

Cluster	Kecamatan
1	Batu Ceper, Karawaci, Ciledug, Pinang.
2	Karang Tengah, Larangan

3 Benda, Cibodas, Cipondoh, Jatiuwung, Neglasari, Periuk, Tangerang.

Tabel 3 merupakan hasil dari pengelompokan menggunakan K-Means Clustering dengan perhitungan Euclidean Distance dan Davies Bouldin Index sebagai pengevaluasi.

4. KESIMPULAN

Dari analisis dan percobaan dapat disimpulkan bahwa Analisis Clustering menggunakan algoritma K-Means yang dirancang untuk mengelompokkan data banjir dari BPBD Kota Tangerang dapat mengidentifikasi daerah dengan tingkat kerentanan terbesar. Sistem ini mampu menampilkan hasil dari tiap-tiap proses perhitungan agar dapat mempermudah user untuk menganalisa data yang di proses menggunakan K-Means Clustering. Menurut penelitian, aplikasi data mining masih jauh dari sempurna. Berdasarkan kesimpulan di atas, dapat disimpulkan bahwa saran-saran berikut dapat berguna untuk pengembangan sistem di masa mendatang. Untuk menghindari kesalahan selama alur program, disarankan untuk melakukan pemeriksaan sistem sebelum mengisi informasi, sehingga jika terjadi kesalahan terkait sistem dapat segera diperbaiki. Pemeliharaan data dan sistem harus diamankan dan didokumentasikan secara teratur sehingga perubahan yang dibuat pada sistem dapat dikenali dan kejadian yang tidak diinginkan dapat diantisipasi. Melakukan pengembangan aplikasi mulai dari segi visual penyajian data agar lebih interaktif bagi user dalam penggunaan yang berkelanjutan nantinya.

DAFTAR PUSTAKA

- [1] P. Septiana Ananda *Et Al.*, “Kmeans Clustering Menggunakan Rapidminer Dalam Segmentasi Pelanggan Dengan Evaluasi Davies Bouldin Index Untuk Menentukan Jumlah Cluster Paling Optimal,” *Jurnal Batirsi*, Vol. 6, No. 2, 2023.
- [2] H. U. Rabbani And D. Anubhakti, *Implementasi Data Mining Untuk Prediksi Tangga Lagu Aplikasi Langit Musik*. Telekomunikasi....
- [3] H. Prastiwi, J. Pricilia, And E. Raswir, “Jurnal Informatika Dan Rekayasa Komputer (Jakakom) Implementasi Data Mining Untuk Menentuksn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering.”
- [4] Y. Pratama, H. Hendrawan, E. Rasywir, B. T. Carenina, And D. R. Anggraini, “Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Provinsi Berdasarkan Banyaknya Desa/Kelurahan Dengan Upaya Antisipasi/Mitigasi Bencana Alam,” *Building Of Informatics, Technology And Science (Bits)*, Vol. 4, No. 3, Dec. 2022, Doi: 10.47065/Bits.V4i3.2549.
- [5] W. Sanusi, A. Zaky, D. Besse, And N. Afni, “Analisis Fuzzy C-Means Dan Penerapannya Dalam Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Faktor-Faktor Penyebab Gizi Buruk.”
- [6] M. R. Irianto, A. Maududie, And F. N. Arifin, “Implementation Of K-Means Clustering Method For Trend Analysis Of Thesis Topics (Case Study: Faculty Of Computer Science, University Of Jember),” *Berkala Sainstek*, Vol. 10, No. 4, P. 210, Dec. 2022, Doi: 10.19184/Bst.V10i4.29524.
- [7] F. Tempola, M. Muhammad, And A. Mubarak, “Penggunaan Internet Dikalangan Siswa Sd Di Kota Ternate: Suatu Survey, Penerapan Algoritma Clustering Dan Validasi Dbi Use Of The Internet In The Elementary School Students In Ternate City: A Survey, Implemented Of Clustering Algorithm And Validation Dbi,” Vol. 7, No. 6, 2020, Doi: 10.25126/Jtiik.202072370.
- [8] R. Julianti Hablum, A. Khairan, J. Jati Metro, And K. Ternate Selatan, “Jiko (Jurnal Informatika Dan Komputer) Ternate Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means,” 2019.
- [9] I. Kadek, J. Arta, G. Indrawan, G. R. Dantes, P. Studi, And I. Komputer, “Data Mining Rekomendasi Calon Mahasiswa Berprestasi Di Smk Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution.”
- [10] S. Informatika And A. Polinema, “Evaluasi Kmeans Clustering Pada Preprocessing Sistem Temu Kembali Informasi,” *Siap*, P. 2020.