

PENERAPAN K-MEANS CLUSTERING UNTUK PEMETAAN WILAYAH RAWAN BENCANA ALAM KOTA MALANG

I Kadek Riski Dwi Putra, Ahmad Faisol, Mira Orisa
Teknik Informatika, Institut Teknologi Nasional Malang
Jalan Raya Karanglo km 2 Malang, Indonesia
2018069@scholar.itn.ac.id

ABSTRAK

Bencana alam merupakan fenomena alam yang mengancam keberlangsungan hidup manusia. Pemetaan wilayah rawan bencana yang dilakukan oleh BPBD Kota Malang dapat menjadi salah satu jawaban dari permasalahan tersebut. Namun pada halaman resmi website BPBD Kota Malang belum memadainya pengelompokan data bencana. Dalam ilmu *sciene* dan teknologi telah banyak permasalahan yang dapat diselesaikan oleh disiplin ilmu data mining untuk mengklasterkan data tertentu. Penggunaan Algoritma *K-Means Clustering* salah satu metode analisis data yang bisa digunakan dalam pengelompokkan wilayah berdasarkan karakteristik yang mirip. Tujuan pengelompokan tersebut diharapkan dapat membantu tugas BPBD Kota Malang untuk mengidentifikasi wilayah dengan tingkat resiko yang sama. Berdasarkan hasil pengujian menggunakan metode *K-Means Clustering* dalam pengelompokan wilayah rawan bencana di Kota Malang pada sistem berjalan dengan baik. Pengujian fungsional yang menunjukkan setiap fitur yang sudah dibuat berjalan sebagaimana fungsinya. Hasil perhitungan manual di excel dan dibandingkan dengan perhitungan sistem yang telah dibuat untuk menentukan *cluster* memiliki tingkat keakurasian sebesar 77,19%.

Kata kunci : BPBD, Bencana alam, Pengelompokkan wilayah, K-Means, Cluster

1. PENDAHULUAN

Bencana alam merupakan fenomena alam yang mengancam keberlangsungan hidup manusia. Kondisi itu membuat Indonesia sangat sering dilanda berbagai jenis bencana alam seperti bencana gempa bumi, tsunami, gunung meletus, tanah longsor, banjir bandang, kekeringan, dan cuaca ekstrem. Tak hanya itu, situasi topografi Indonesia dengan banyaknya kawasan daerah pegunungan menjadikan faktor terjadinya bencana alam di berbagai tempat.

Pemetaan wilayah rawan bencana yang dilakukan oleh BPBD Kota Malang dapat menjadi salah satu jawaban dari permasalahan tersebut. Namun pada halaman resmi website BPBD Kota Malang belum memadainya pengelompokan data bencana yang dapat membantu masyarakat untuk mempersiapkan hal dalam menghadapi suatu bencana kedepannya.

Dalam ilmu *sciene* dan teknologi telah banyak permasalahan yang dapat diselesaikan oleh disiplin ilmu data mining untuk mengklasterkan data tertentu. penggunaan Algoritma *K-Means Clustering* yang merupakan salah satu metode analisis data yang bisa digunakan dalam pengelompokkan wilayah berdasarkan karakteristik yang mirip. Dengan penerapan metode ini, diharapkan wilayah yang memiliki karakteristik serupa dalam hal risiko bencana alam dapat dikelompokkan dalam satu kelompok. Tujuan pengelompokan tersebut diharapkan dapat membantu tugas BPBD Kota Malang untuk mengidentifikasi wilayah dengan tingkat resiko yang sama untuk segera memberikan bantuan ataupun penanggulangan masalah yang terjadi. Pembagian karakteristik kelompok bencana

alam berdasarkan tingkat kawasannya pun dapat dibantu melalui metode ini.

Pada penelitian ini penerapan algoritma *K-Means Clustering* dalam pemetaan daerah rawan bencana di Kota Malang dapat membantu dalam pengembangan sistem informasi yang lebih baik khususnya bagi website BPBD Kota Malang. Dengan sistem ini, pemetaan dalam kelompok wilayah yang rawan bencana akan dapat diidentifikasi lebih mudah serta penanganan dapat dilakukan dengan lebih cepat dan tepat. Hal ini sangat penting dalam mengurangi dampak bencana alam terhadap penduduk serta menjadi acuan dalam penentuan langkah pengambilan keputusan di kemudian hari nantinya. Dengan adanya penerapan algoritma *K-Means Clustering* di Kota Malang, diharapkan dapat meningkatkan ketangguhan komunitas terhadap bencana alam.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian yang dilakukan oleh Rif'atul Amaliah dan rekannya yang berjudul "Pengelompokkan Data Bencana Alam berdasarkan Wilayah Menggunakan Algoritma K-Means", dengan tujuan yaitu memberikan hasil informasi dalam bentuk pola pengelompokkan daerah rawan bencana mulai dari tahun 2021-2022 yang akan dijadikan sebagai acuan di dalam pengambilan keputusan pihak-pihak terkait untuk mengidentifikasi wilayah-wilayah dimana jenis bencana alam paling dominan atau sering terjadi.[1]

Menurut Tutut Suryani, Ahmad Faisol, dan Nurlaily Vendyansyah dalam penelitiannya yang berjudul "Sistem Informasi Geografis Pemetaan

Kerusakan Jalan di Kota Malang Menggunakan Metode K-Means” yang bertujuan untuk membuat sistem informasi geografis pemetaan kerusakan jalan oleh warga yang diharapkan Pemerintah dapat mengelompokkan tingkat kerusakan jalan di Kabupaten Malang menggunakan metode k-means. Metode kmeans adalah metode yang dapat digunakan untuk mengelompokkan tingkat kerusakan jalan sehingga pemerintah mengetahui tingkatan rusak berat dapat ditangani terlebih dahulu. [2]

Penelitian yang dilakukan oleh Edi Wahyudin, Rizki Amir Rudin, Kaslani, dan Sandy Eka Permana yang berjudul “Penerapan Data Mining Pengelompokan Produktivitas Padi Menggunakan Algoritma K-Means Pada Provinsi Jawa Barat”, bertujuan untuk memberikan wawasan terkait daerah. Informasi ini tidak hanya bermanfaat bagi pemerintah dalam perumusan kebijakan pertanian yang efektif, tetapi juga sebagai landasan bagi penelitian masa depan yang memperhatikan potensi produktivitas padi [3].

2.2. Data Mining

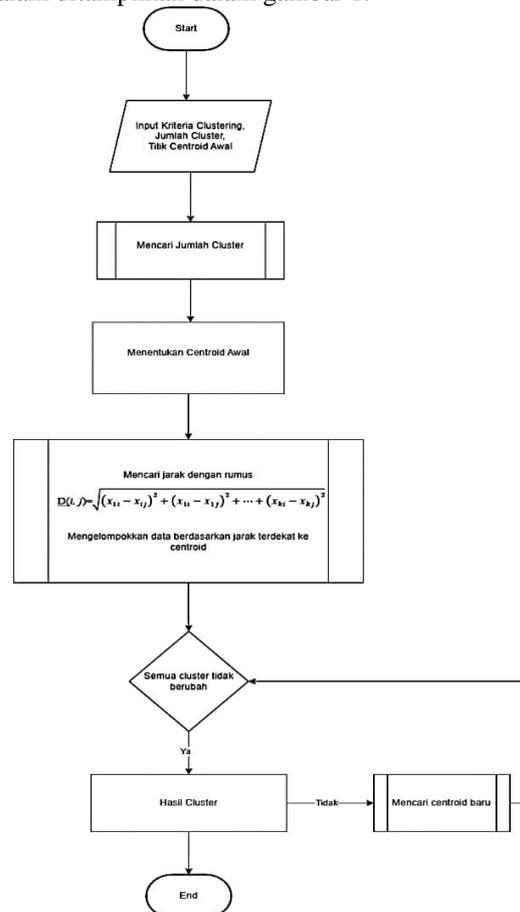
Data mining merupakan salah satu ilmu bidang komputer yang telah banyak digunakan untuk menyelesaikan berbagai kasus penelitian. Terdapat berbagai macam algoritma dalam data mining seperti K-Means, K-Medoids, C4.5, Naïve Bayes, Regresi Linier dan lainnya. Adapun istilah yang ada pada pada kata data mining yaitu seperti pattern recognition yang maksudnya yaitu data mining dapat melakukan pengenalan pola data yang bersifat tersembunyi pada suatu bongkahan data. Ada pula istilah knowledge discovery yang maksudnya data mining dapat dilakukan penemuan pengetahuan pada data tersembunyi yang terdapat di dalam bongkahan data dikarenakan pada realitanya data mining digunakan untuk memperoleh pengetahuan secara terperinci.[4]

Secara luas, pengertian data mining adalah proses mengekstrak informasi yang berguna dari data yang besar dan kompleks. Ini mencakup teknik statistik, algoritma, dan proses yang digunakan untuk menemukan pola dan hubungan dalam data. Tujuan dari data mining adalah untuk menemukan informasi yang berguna dan mengubahnya menjadi bentuk yang dapat diterima oleh pengguna. Data mining digunakan dalam berbagai bidang, seperti bisnis, ilmu pengetahuan, teknologi informasi, dan sebagainya untuk menemukan pola dalam data yang dapat digunakan untuk membuat keputusan yang lebih baik.[5]

2.3. Algoritma K-Means

Metode K-means merupakan salah satu strategi pengelompokan data nonhierarki yang bertujuan mempartisi data ke dalam kelompok-kelompok. Prinsip dasarnya adalah mengelompokkan data sedemikian rupa sehingga data dengan karakteristik serupa ditempatkan dalam satu kelompok, sementara data yang memiliki karakteristik berbeda ditempatkan dalam kelompok lain.[6]

Diagram *flowchart* metode pada aplikasi pemetaan daerah yang rentan terhadap bencana alam ditampilkan dalam gambar 1.



Gambar 1. Flowchart Metode K-Means Clustering

Pada diagram 1, terdapat diagram *flowchart* yang menjelaskan proses dari algoritma *K-Means Clustering*. Langkah awal melibatkan masukan kriteria *clustering*, jumlah *cluster*, dan penentuan titik awal *centroid* yang akan digunakan dalam perhitungan. Setelah itu, proses melibatkan penentuan jumlah *cluster* dan penentuan titik awal *centroid*. Selanjutnya, data dikelompokkan berdasarkan jarak terdekat ke *centroid* dengan menggunakan perhitungan sesuai rumus yang telah ditetapkan. Jika tidak ada perubahan pada seluruh *cluster*, hasil *cluster* akan ditentukan. Namun, jika ada data yang berubah, proses akan kembali mencari *centroid* baru.

Langkah-langkah untuk menjalankan proses pengelompokan dengan metode *K-Means* adalah sebagai berikut:

1. Menentukan jumlah *cluster* (k) yang diinginkan.
2. Melakukan inisialisasi k pusat *cluster*. Inisialisasi ini dapat dilakukan dengan berbagai metode, namun pilihan umumnya adalah penggunaan nilai acak. Pusat-pusat *cluster* diberikan nilai awal yang dihasilkan secara acak.

- Mengalokasikan semua data atau objek ke dalam *cluster* yang paling dekat. Kedekatan dua objek ditentukan berdasarkan jarak di antara keduanya. Dalam tahap ini, perlu dihitung jarak setiap data ke setiap pusat *cluster*. Jarak antara suatu data dengan suatu *cluster* tertentu akan menentukan keanggotaan data tersebut dalam *cluster* tersebut. Untuk menghitung jarak, dapat menggunakan teori jarak *Euclidean*, yang dirumuskan sebagai berikut:

$$D(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ki} - x_{kj})^2} \dots\dots (1)$$

Dimana :

$D(x_i, x_j)$ = Jarak dari data ke i ke pusat *cluster* j

x_{ki} = Data ke i pada atribut ke k

x_{kj} = Data ke j pada atribut ke k

- Perbaharui pusat *cluster* dengan memperhitungkan keanggotaan *cluster* saat ini. Pusat *cluster* dihitung sebagai rata-rata dari data atau objek yang termasuk dalam *cluster* tertentu. Jika diinginkan, median dari *cluster* tersebut juga dapat digunakan sebagai alternatif ukuran. Oleh karena itu, penggunaan rata-rata (*mean*) tidaklah satu-satunya metode yang dapat digunakan
- Perbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dengan menghitung rata-rata dari *cluster* yang bersangkutan, menggunakan rumus yang telah ditentukan :

$$C_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_{i(q)} \dots\dots\dots (2)$$

Dimana:

$C_{m(q)}$ = Centroid dari klaster m pada iterasi ke-q

N_m = Jumlah data yang termasuk dalam klaster m

$X_{i(q)}$ = Data ke-i pada iterasi ke-q

- Lakukan proses pada setiap objek dengan menggunakan pusat *cluster* yang telah diperbarui. Jika pusat *cluster* tidak mengalami perubahan, maka proses pengelompokan selesai. Namun, jika terdapat perubahan, kembali ke langkah nomor 3 dan teruskan proses tersebut hingga pusat *cluster* tidak berubah lagi

2.4. Website

Dalam istilah penggunaan, sebuah situs *web* merupakan koleksi halaman yang umumnya tergabung dalam satu domain atau subdomain, yang berada di dalam *World Wide Web (WWW)* di Internet. Beberapa situs web mengharuskan pengguna untuk berlangganan (memberikan data masukan) agar dapat mengakses sebagian atau seluruh konten yang tersedia. Sebagai contoh, terdapat beberapa situs bisnis dan layanan surel gratis yang memerlukan langganan agar pengguna dapat mengakses situs tersebut [7].

2.5. Database

Database dipergunakan untuk menyimpan data yang terkait dengan suatu sistem atau organisasi, dan memfasilitasi pengguna dalam menarik data sesuai dengan keperluan mereka. Jenis-jenis database juga

dapat dikelompokkan, seperti *database* relasional, database berorientasi objek, dan database terdistribusi, bergantung pada cara penyimpanan dan organisasi data dalam sistem[8].

2.6. PHP

PHP merupakan bahasa skrip yang ditanam atau disisipkan ke dalam *HTML*. Untuk memisahkan antara kode *PHP* dan kode *HTML* sebagai tempatnya, tag-tag *PHP* digunakan. *PHP* memiliki popularitas tinggi dan dapat digunakan untuk mengembangkan berbagai jenis situs *web* dinamis, bahkan dapat digunakan untuk membuat sistem manajemen konten (*CMS*) [9].

2.7. Laravel

Laravel adalah kerangka kerja yang mendukung pengembang *web* dalam mengoptimalkan penggunaan *PHP* pada tahapan pembangunan situs web.

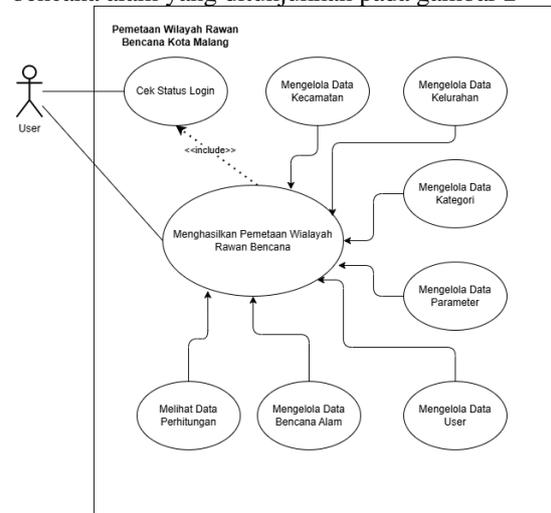
Laravel memungkinkan membuat seluruh proses pengembangan *web* lebih cepat dan lebih mudah bagi pengembang dengan menghilangkan semua masalah yang terkait dengan penanganan kode *PHP* yang rumit. [10]

3. METODE PENELITIAN

Dalam suatu studi, perlu memberikan perhatian khusus pada pemilihan metode penelitian agar dapat mencapai hasil yang optimal. Penelitian ini memanfaatkan metode *K-Means Clustering* Algorithm untuk melakukan pengelompokan wilayah rawan bencana alam di Kota Malang

3.1. Use Case Diagram

Use Case website pengelompokan wilayah rawan bencana alam yang ditunjukkan pada gambar 2



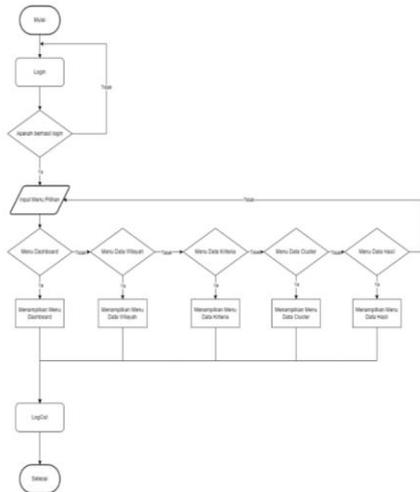
Gambar 2. Use Case Diagram

Pada ilustrasi dalam Gambar 2 terdapat satu pengguna, yaitu *administrator*. *Administrator* memiliki tanggung jawab untuk mengelola setiap data yang diperlukan oleh sistem untuk melakukan pengelompokan dengan menggunakan Algoritma *K-Means Clustering*. *Administrator* memiliki kemampuan untuk menambah, mengedit, dan menghapus berbagai jenis data yang diperlukan

dalam sistem, seperti data kategori, parameter, informasi bencana, dan proses klasterisasi.

3.2. Flowchart Sistem

Flowchart sistem pada website pengelompokan wilayah rawan bencana alam yang ditunjukkan pada gambar 3

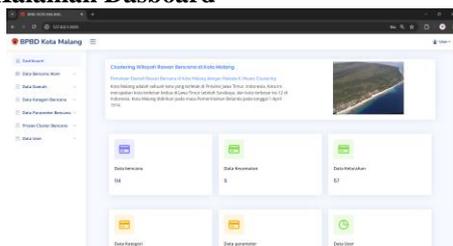


Gambar 3. Flowchart sistem

Pada Gambar 3, terlihat diagram alir yang menggambarkan langkah-langkah penggunaan sistem. Admin diarahkan ke halaman login, dan jika berhasil login, akan diteruskan ke halaman dashboard. Jika login tidak berhasil, admin akan kembali ke halaman login. Terdapat berbagai menu, termasuk menu untuk melakukan operasi *CRUD* pada data nama kecamatan yang terdampak, nama kelurahan yang terdampak, kategori bencana, data parameter, data bencana, dan proses *clustering*. Menu proses clustering digunakan untuk melaksanakan perhitungan pengelompokan wilayah rawan bencana dengan menggunakan metode *K-Means*.

4. HASIL DAN PEMBAHASAN

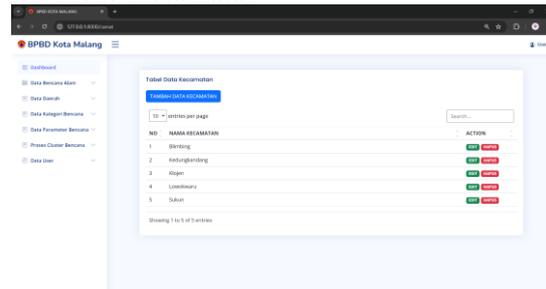
4.1. Halaman Dashboard



Gambar 4. Halaman dashboard

Halaman dashboard merupakan tampilan halaman dashboard yang sudah dibuat. Pada halaman dashboard terdapat profile Kota Malang, menu data bencana, data kecamatan, data kelurahan, data parameter, data kategori dan data user.

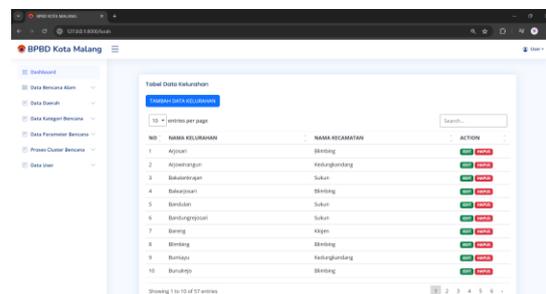
4.2. Menu Kecamatan



Gambar 5. Halaman Data Kelurahan

Halaman data kecamatan adalah antarmuka yang menampilkan informasi mengenai kecamatan yang telah dibuat. Dalam tabel kategori, terdapat kolom-kolom seperti nomor, nama kecamatan, dan informasi terkait dampak yang terjadi

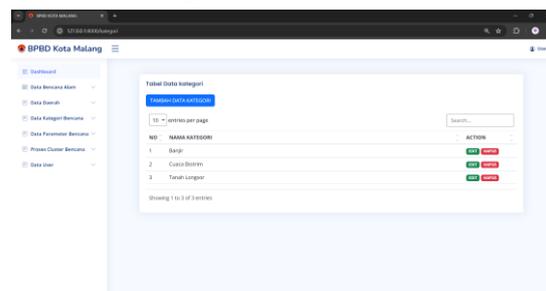
4.3. Menu Kelurahan



Gambar 6. Halaman Data Kelurahan

Halaman data kelurahan adalah antarmuka yang menampilkan informasi mengenai data kelurahan yang telah dibuat. Dalam tabel kategori, terdapat kolom-kolom seperti nomor, nama kelurahan, dan nama kecamatan.

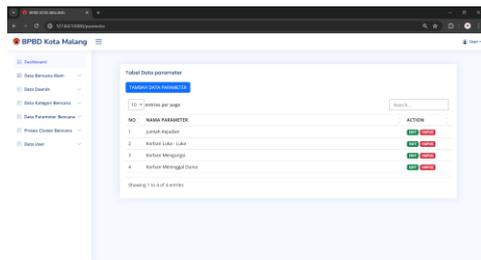
4.4. Menu Kategori



Gambar 7. Halaman Kategori

Halaman kategori merupakan tampilan halaman kategori yang sudah dibuat. Pada tabel kategori terdapat field no, nama kategori bencana yang terjadi

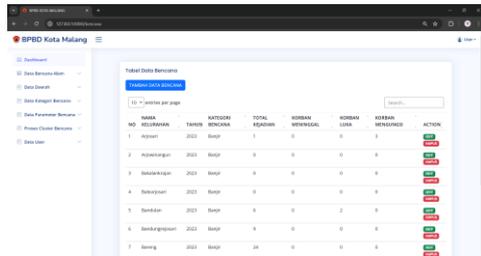
4.5. Menu Parameter



Gambar 8. Halaman Parameter

Halaman parameter merupakan tampilan halaman parameter yang sudah dibuat. Pada table parameter terdapat yaitu *field* id dan nama parameter yang disebabkan oleh bencana alam.

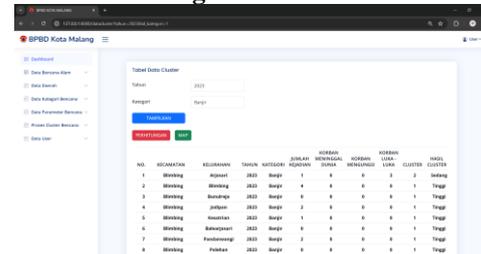
4.3 Data Bencana



Gambar 9. Halaman Data Bencana

Halaman data bencana merupakan tampilan data bencana yang sudah di buat. Halaman ini berisi *field* dari nama wilayah, tahun, parameter dan jenis bencana

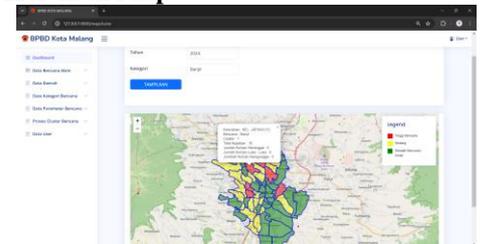
4.4 Proses Clustering



Gambar 10. Halaman Proses Clustering

Halaman proses *clustering* adalah antarmuka yang menampilkan data yang akan dikelompokkan. Halaman ini memuat berbagai *field*, termasuk nama wilayah, tahun, kategori, parameter, *cluster* yang terbentuk, dan jenis klaster yang diterapkan.

4.5 Halaman Map



Gambar 11. Halaman Proses Clustering

Gambar 11 adalah halaman map yang akan berisikan kelompok yang sudah di bagi menjadi beberapa *cluster* dan diberi warna *cluster* 1 merah, *cluster* 2 kuning, dan *cluster* 3 hijau

4.6. Implementasi Metode K-Means Clustering

Tabel 1. Input data wilayah

No	Kelurahan	Kecamatan
1	Arjosari	Blimbing
2	Arjowinangun	Kedungkandang
...
57	Wonokoyo	Kedungkandang

Untuk menghitung nilai proses iterasi, tahap awal melibatkan penjumlahan nilai K-n dengan menggunakan rumus tertentu. Rumus tersebut mengacu pada bobot terjadinya bencana di suatu daerah, jumlah korban yang meninggal, jumlah korban luka dan jumlah korban mengungsi yang telah diidentifikasi dengan kode a, b, c, d. Nilai-nilai ini telah diinputkan sebelumnya dalam sistem:

Tabel 2. Parameter bencana

No	(A)	(B)	(C)	(D)
1	0	0	0	0
2	1	0	0	0
...
57	0	0	0	0

Penjelasan parameter diatas:

1. Jumlah Kejadian (A): Jumlah kejadian bencana mengacu pada jumlah kejadian bencana yang terjadi pada setiap kelurahan di Kota Malang
2. Korban Meninggal (B): Korban Meninggal merujuk pada jumlah korban yang meninggal diakibatkan yang ada pada setiap kelurahan.
3. Korban Luka (C): Korban Luka merujuk pada jumlah korban yang terluka diakibatkan bencana alam pada setiap kelurahan.
4. Korban Mengungsi (D): Korban Mengungsi merujuk pada jumlah korban yang mengungsi diakibatkan bencana alam pada setiap kelurahan

Berikut merupakan implementasi perhitungan metode *K-Means* pada aplikasi:

1. Menentukan jumlah *cluster* dan menentukan pusat *cluster / centroid* secara acak

```

1 $k = 3;
2 $lengthData = count($this->dataBencana_old);
3 $centroid_data = [];
4 $c1 = Session::get('c1');
5 $c2 = Session::get('c2');
6 $c3 = Session::get('c3');
    
```

Gambar 11. Source code menentukan jumlah *cluster*

Pada gambar 4.10 merupakan source code yang digunakan untuk menentukan titik *centroid* awal.

2. Menentukan Jarak

```

1 $hitungC[$j] = sqrt(pow(($this->dataBencana_old[$i]['data']->nilai_1
2 $dataCentroid[$j]['nilai_1']), 2) +
3 pow(($this->dataBencana_old[$i]['data']->nilai_2
4 - $dataCentroid[$j]['nilai_2']), 2) +
5 pow(($this->dataBencana_old[$i]['data']->nilai_3
6 - $dataCentroid[$j]['nilai_3']), 2) +
7 pow(($this->dataBencana_old[$i]['data']->nilai_4
8 - $dataCentroid[$j]['nilai_4']), 2));
9
    
```

Gambar 12. Source code perhitungan Euclidean Distance

Pada gambar selanjutnya akan di dapat nilai *centroid* baru untuk acuan perhitungan berikutnya sampai nilai *centroid* sebelum dan sesudah bernilai sama. Selanjutnya menentukan jarak *centroid* atau nilai minimum. Klasifikasi setiap data berdasarkan kedekatannya dengan *centroid*

3. Memperbaharui Cluster

```

1 $centroid_data = [];
2 for ($i = 0; $i < count($dataCentroid); $i++) {
3     $nilai_1 = 0;
4     $nilai_2 = 0;
5     $nilai_3 = 0;
6     $nilai_4 = 0;
7     $jumlahData = 0;
8     for ($j = 0; $j < count($dataBaru); $j++) {
9         // dd($dataBaru[$j]['cluster']);
10        if ($dataBaru[$j]['cluster'] == ($i + 1)) {
11            // $centroid_data[$i][$j] = $dataBaru[$j];
12            $nilai_1 += $dataBaru[$j]['data']->nilai_1;
13            $nilai_2 += $dataBaru[$j]['data']->nilai_2;
14            $nilai_3 += $dataBaru[$j]['data']->nilai_3;
15            $nilai_4 += $dataBaru[$j]['data']->nilai_4;
16            $jumlahData ++ 1;
17        }
18    }

```

Gambar 13. Source code memperbaharui data cluster Pada gambar 4.12 setelah klasifikasi setiap data berdasarkan kedekatannya dengan *centroid*, selanjutnya memperbaharui data cluster lalu dilanjutkan kembali dengan perhitungan jarak dengan *centroid* dengan nilai c sampai anggota tiap cluster tidak ada yang berubah. Jika anggota tiap cluster tidak berubah, maka nilai pusat cluster pada iterasi terakhir akan digunakan sebagai parameter untuk menentukan klasifikasi data.

Perhitungan Manual K-Means

1. Menentukan jumlah cluster

Tabel 3. Tabel cluster

Cluster	Nama Cluster
C1	Tinggi
C2	Sedang
C3	Rendah

2. Menentukan pusat cluster (*centroid*) = acak

Pada perhitungan excel, nilai k yang digunakan sebanyak 4. Setelah menentukan jumlah k, pada tahap kedua yaitu memilih pusat cluster secara acak yang terdiri dari data ke 5, 8, dan 2

Tabel 4. Tabel Centroid

Data ke	centroid	A	B	C	D
5	C1	5	0	4	0
8	C2	2	0	0	0
2	C3	1	0	0	0

3. Menentukan jarak

Setelah mendapatkan titik pusat awal klaster, langkah selanjutnya adalah melakukan perhitungan jarak *Euclidean* dan mengelompokkan data berdasarkan jarak terpendek. Kemudian, nilai *centroid* baru akan dihasilkan untuk menjadi acuan dalam perhitungan berikutnya, terus-menerus dilakukan proses ini hingga nilai *centroid* sebelum dan sesudahnya menjadi sama.

$$D(x_i, x_j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ki} - x_{kj})^2}$$

... (3)

1) Berikut ini beberapa hasil perhitungan jarak *Euclidean* pada titik pusat 1

$$D(1,1) = \sqrt{(0 - 5)^2 + (0 - 0)^2 + (0 - 4)^2 + (0 - 0)^2} = 6,403124237$$

$$D(57,1) = \sqrt{(0 - 5)^2 + (0 - 0)^2 + (0 - 4)^2 + (0 - 0)^2} = 6.403124237$$

2) Berikut ini beberapa hasil perhitungan jarak *Euclidean* pada titik pusat 2

$$D(2,0,0,0) = \sqrt{(0 - 2)^2 + (1 - 0)^2 + (0 - 0)^2 + (1 - 0)^2} = 2$$

$$D(57,2) = \sqrt{(0 - 2)^2 + (1 - 0)^2 + (0 - 0)^2 + (1 - 0)^2} = 2$$

3) Berikut beberapa hasil perhitungan jarak *Euclidean* pada pusat titik 3.

$$D(1,0,0,0) = \sqrt{(0 - 1)^2 + (0 - 0)^2 + (0 - 0)^2 + (1 - 0)^2} = 1$$

$$D(57,3) = \sqrt{(0 - 1)^2 + (0 - 0)^2 + (1 - 0)^2 + (0 - 0)^2} = 0$$

Setelah melakukan perhitungan untuk seluruh data dengan menggunakan rumus *Euclidean*, langkah berikutnya adalah mengelompokkan data ke dalam klaster berdasarkan nilai terkecil atau jarak terpendek ke masing-masing klaster. Setelah tahap pengelompokkan selesai, langkah selanjutnya adalah menghitung nilai *centroid* baru dengan memanfaatkan persamaan 1.

4. Menentukan Jarak Centroid (Nilai Minimum)

Mengelompokkan data objek berdasarkan jarak minimum ke *centroid*. Dimana, C1 dikategorikan sebagai cluster 1 (Tinggi), C2 sebagai cluster 2 (Sedang), dan C3 sebagai cluster 3 (Rendah/Tidak Beresiko)

No	Kelurahan	Centroid			Min	C
		C1	C2	C3		
1	Arjosari	6.403 12423 7	2	1	1	3
2	Arjowinangun	5.656 85424 9	1	0	0	3
...		
57	Wonokoyo	6.403 12423 7	2	1	1	3

5. Pengelompokan setiap data berdasarkan kedekatannya dengan *centroid*.

Tabel 6 Klasifikasi setiap data dengan *centroid*

No	Nama Kelurahan	C1	C2	C3
1	Arjosari			1
2	Arjowinangun			1
...
57	Wonokoyo			1

6. Memperbaharui Cluster

Memeriksa apakah terdapat objek yang mengalami perubahan kelompok atau tidak.

Apabila iya, proses perulangan akan dilakukan dari tahap kedua hingga tahap kelima. Apabila tahap terakhir terpenuhi, nilai pusat klaster pada iterasi terakhir akan dijadikan sebagai parameter untuk menentukan klasifikasi data. Selanjutnya, dilakukan pembaruan nilai *centroid* dengan menghitung rata-rata klaster yang bersangkutan menggunakan rumus:

$$C_{m(q)} = \frac{1}{n_m} \sum_{i=1}^{n_m} x_{i(q)} \dots\dots\dots (4)$$

Tabel 7 Klasifikasi Berdasarkan Kedekatan Dengan *Centroid*

No	Nama Kelurahan	C1	C2	C3
1	Arjosari			1
2	Arjowinangun			1
...
57	Wonokoyo			1

Setelah memperoleh nilai *centroid* yang baru, lanjutkan dengan menghitung nilai jarak *Euclidean* untuk setiap data terhadap *centroid* baru pada iterasi kedua. Selanjutnya, lakukan perhitungan *centroid* baru secara berulang hingga nilai *centroid* baru tidak mengalami perubahan dari nilai *centroid* lama atau iterasi sebelumnya. Berikut merupakan table hasil pengolahan data yang dilakukan ditampilkan pada tabel 8

Tabel 8 Hasil *Clustering*

No	C1	C2	C3	Minimum	Cluster
1.	6.94 4222 2	2.61 9905 6	0.46 0194 2	0.460194202	Cluster3
2.	5.99 0733 6	1.62 2694 1	0.65 6727 4	0.656727416	Cluster3
....
57.	6.94 4222 2	2.61 9905 6	0.46 0194 2	0.460194202	Cluster3

4.7. Pengujian Black Box

Pemisahan setiap data berdasarkan kedekatannya dengan *centroid*.

Tabel 10 Pengujian Black-Box

Menu	Aktivitas pengujian	Hasil diharapkan	Hasil pengujian	Kesimpulan
Login	Memasukkan <i>password</i> dan <i>user</i> benar	<i>Email</i> = "admin@gmail.com" <i>Password</i> = "12345678"	<i>System</i> menerima akses melihat <i>role</i>	Valid
Data User	Memasukkan data user	Nama = "Cahyanto" Email = "cahyanto@gmail.com" Password = "12345678"	Sistem menerima direct ke halaman data user	valid
Data Kecamatan	Memasukkan Kecamatan	Kecamatan = "Blimbing"	Sistem menerima direct ke halaman	valid

Data Kelurahan	Memasukkan kelurahan	Kelurahan = "Arjosari"	Sistem menerima direct ke halaman data kelurahan	valid
Data Kategori	Memasukkan kategori	Kategori = "Banjir"	Sistem menerima direct ke halaman data kategori	valid
Data Parameter	Memasukkan parameter	Parameter = "Kerugian Fisik"	Sistem menerima direct ke halaman data parameter	valid
Data Bencana	Memasukkan data bencana	Kelurahan = "Arjosari" Tahun="2023" Parameter = "Kejadian Bencana" Nilai parameter = "2" Kategori = "Banjir"	Sistem menerima direct ke halaman data bencana	valid
Map	Memilih kategori bencana dan tahun yang diinginkan	Tahun = "2023" Bencana = "Banjir"	Sistem menampilkan bencana yang dipilih sesuai dengan kategori dan tahun	Valid
Logout	Melakukan <i>logout</i>	Melakukan <i>logout</i>	Sistem <i>direct</i> kehalaman <i>login</i>	Valid

4.8. Pengujian Metode *K-Means Clustering*

Pengujian ini dilaksanakan untuk menilai tingkat keakuratan metode *K-Means Clustering*. Dalam pengujian ini, peneliti menggunakan 57 data terkait Cuaca Ekstrem pada tahun 2022. Berikut merupakan gambar dari hasil *clustering* pengelompokan wilayah rawan bencana alam pada sistem

Tabel 12 Data Asli

No	Nama Kelurahan	A	B	C	D	Kelas
5	Kesatrian	2	0	0	0	Cluster 2
14	Buring	5	0	4	0	Cluster 3
16	Kedungkandang	3	0	1	0	Cluster 2
21	Sawojajar	2	0	1	0	Cluster 2
25	Gadingasri	6	0	1	0	Cluster 1
26	Kasin	2	0	0	0	Cluster 2
27	Kauman	3	0	0	0	Cluster 2
29	Klojen	1	0	0	8	Cluster 3
30	Oro – oro dowo	9	0	0	3	Cluster 1
40	Mojolangu	3	0	0	0	Cluster 2
41	Sumbersari	2	0	0	0	Cluster 2
44	Tulusrejo	2	0	0	0	Cluster 2
46	Tunjung sekar	2	0	0	0	Cluster 2

Pada tabel 12 Terdapat perbedaan data antara data actual dengan hasil clustering sistem. Perbedaan ini terletak pada kelurahan buring, gading asri, kasin kauman, kedungkandang, kesantrian, mojolangu, oro-oro dowo, sawojajar, sumbersari, tulusrejo, tunjung sekar, dan klojen

5	Blimbing	Kesantrian	2023	Cuaca Ekstrem	2	0	0	0	3	Rendah
14	Kedungkandang	Buring	2023	Cuaca Ekstrem	5	0	4	0	2	Sedang
16	Kedungkandang	Kedungkandang	2023	Cuaca Ekstrem	3	0	1	0	3	Rendah
21	Kedungkandang	Sawojajar	2023	Cuaca Ekstrem	2	0	1	0	3	Rendah
25	Klojen	Gadingasri	2023	Cuaca Ekstrem	6	0	1	0	2	Sedang
26	Klojen	Kasin	2023	Cuaca Ekstrem	2	0	0	0	3	Rendah
27	Klojen	Kauman	2023	Cuaca Ekstrem	3	0	0	0	3	Rendah
29	Klojen	Klojen	2023	Cuaca Ekstrem	1	0	0	8	1	Tinggi
30	Klojen	Oro-Oro Dowo	2023	Cuaca Ekstrem	9	0	0	3	2	Sedang
40	Lowokwaru	Mojolangu	2023	Cuaca Ekstrem	3	0	0	0	3	Rendah
41	Lowokwaru	Sumbersari	2023	Cuaca Ekstrem	2	0	0	0	3	Rendah
44	Lowokwaru	Tulusrejo	2023	Cuaca Ekstrem	2	0	0	0	3	Rendah
46	Lowokwaru	Tunjungsekar	2023	Cuaca Ekstrem	2	0	0	0	3	Rendah

Maka berdasarkan perbandingan antara data hasil clustering dari sistem dan data aktual dari lapangan yang telah di lakukan, kemudian akan dihitung presentase akurasi sesuai dengan Persamaan 5.

$$\text{Akurasi} = \frac{\text{Jumlah Profil Yang Sesuai}}{\text{Total Jumlah Profil}} \times 100\% \dots (5)$$

Pada tahun 2022 telah dilakukan pengujian terhadap 57 data, dari pengujian yang telah dilakukan terdapat data yang akurat sebanyak 57 data. Dengan hasil data tersebut dapat menghasilkan nilai akurasi sesuai dengan Persamaan 6.

$$\text{Akurasi} = \frac{44}{57} \times 100\% = 77,19\% \dots (6)$$

5. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil pengujian menggunakan metode *K-Means Clustering* dalam pengelompokan wilayah rawan bencana di Kota Malang pada sistem berjalan dengan baik. Pengujian fungsional yang menunjukkan setiap fitur yang sudah dibuat berjalan sebagaimana fungsinya. Hasil perhitungan manual di *excel* dan dibandingkan dengan perhitungan sistem yang telah dibuat untuk menentukan *cluster* memiliki tingkat keakurasian sebesar 77,19%.

5.2. Saran

Peningkatan dan pengembangan pada klasifikasi wilayah rawan bencana alam dengan menggunakan Algoritma *K-Means Clustering* dapat diaplikasikan dengan menyelaraskannya ke dalam aplikasi *Android*. Tujuan utamanya adalah untuk meningkatkan efisiensi akses pengguna, memfasilitasi mereka untuk menggunakan *algoritma* tersebut secara langsung melalui perangkat *smartphone* tanpa bergantung pada aplikasi *browser*. dan menambahkan perhitungan untuk pemilihan optimasi pusat centroid.

DAFTAR PUSTAKA

[1] R. Amaliah, E. Tohidi, E. Wahyudin, A. Rizki Rinaldi, and I. Iin, "Pengelompokan Data Bencana Alam Berdasarkan Wilayah Menggunakan Algoritma K-Means," *JATI*

(*Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3572–3579, 2024, doi: 10.36040/jati.v7i6.8253.

- [2] T. Suryani, A. Faisol, and N. Vendyansyah, "Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Kabupaten Malang Menggunakan Metode K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 5, no. 1, pp. 380–388, 2021, doi: 10.36040/jati.v5i1.3259.
- [3] E. Wahyudin, R. Amir Rudin, and S. Eka Permana, "Penerapan Data Mining Pengelompokan Produktivitas Padi Menggunakan Algoritma K-Means Pada Provinsi Jawa Barat," *J. Mhs. Tek. Inform.*, vol. 8, no. 1, 2024, [Online]. Available: <https://opendata.jabarprov.go.id/id/dataset/produktivitas>
- [4] M. Rafi Nahjan, Nono Heryana, and Apriade Voutama, "Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 101–104, 2023, doi: 10.36040/jati.v7i1.6094.
- [5] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *J-SISKO TECH (Jurnal Teknol. Sist. Inf. dan Sist. Komput. TGD)*, vol. 6, no. 1, p. 164, 2023, doi: 10.53513/jsk.v6i1.7351.
- [6] Nurjoko, D. Dwirohayati, and N. H. Sudibyo, "Sistem Informasi Pemetaan Wilayah Rawan Kriminalitas Polresta Bandar Lampung Menggunakan K-Means Clustering," *Teknika*, vol. 14, no. 2, pp. 127–135, 2020.
- [7] . N., A. Ibrahim, and A. Ambarita, "Sistem Informasi Pengaduan Pelanggan Air Berbasis Website Pada Pdam Kota Ternate," *IJIS - Indones. J. Inf. Syst.*, vol. 3, no. 1, p. 10, 2018, doi: 10.36549/ijis.v3i1.37.
- [8] N. Novita, "Manajemen Proyek Sistem Informasi Pengolahan Data Apotek Berbasis Database," *Methosisfo J. Ilm. Sist. Inf.*, vol. 2, no. 1, pp. 9–17, 2022.
- [9] N. Rubiati, "Aplikasi Informasi Pelayanan Fitness Pada Golden Fitness Center Dumai Dengan Bahasa Pemrograman Php," *INF O R M a T I K a*, vol. 10, no. 1, p. 1, 2018, doi: 10.36723/juri.v10i1.53.
- [10] S. Arfan and S. I. G. Novian, "Implementasi Aplikasi Framework Laravel PT XYZ," *J. Tek. Inform. Stmik Antar Bangsa*, vol. V, no. 1, pp. 18–24, 2019.