

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Data yang digunakan dalam penelitian ini didapat dari <https://opendata.jabarprov.go.id/id/dataset> dengan sumber dari Badan Penanggulangan Bencana Daerah (BPBD) yang diakses pada pukul 21:23 tanggal 12 Oktober 2022 yaitu data Jumlah Kerusakan Rumah Tempat Tinggal Akibat Bencana yang akan di *cluster* berdasarkan Kabupaten atau Kota mana saja yang rawan akan bencana dengan kerusakan paling beresiko di Jawa Barat pada 10 tahun terakhir yaitu 2012-2021 yang berjumlah 1620 data, dimana data tersebut berisi nama Kabupaten atau Kota, kondisi kerusakan, jumlah kerusakan.

3.2. Peralatan Penelitian

Penelitian ini menggunakan berupa perangkat keras dan lunak yaitu :

1. Perangkat Keras (*Hardware*)
 - a. Laptop ASUS AMD A6 dengan spesifikasi sebagai berikut .
 - Sistem Operasi Windows 10
 - RAM 4GB
 - HDD 1TB
 - X441B
 2. Software yang digunakan pada laptop antara lain :
 - a. Website Opendata.jabarprov (sebagai media pengambilan data)
 - b. Microsoft Office Word 2019
 - c. Microsoft Office Excel
 - d. Python

3.3 Lokasi dan Waktu Penelitian

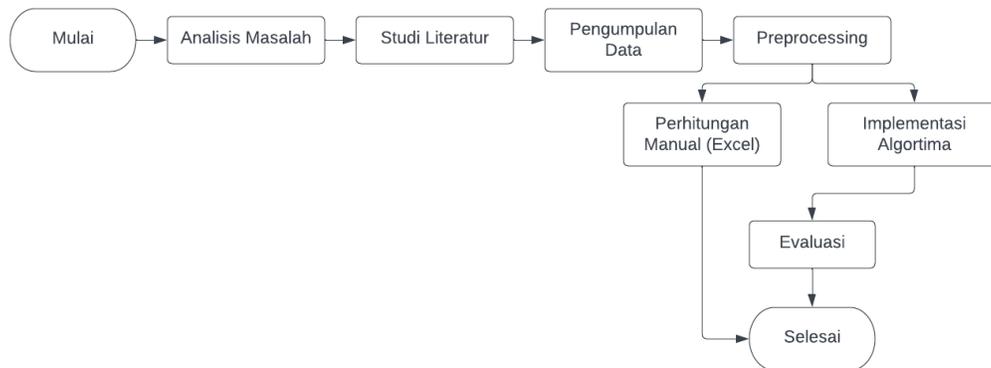
Penelitian dilakukan di Universitas Buana Perjuangan Karawang. Kemudian pelaksanaan penelitian berlangsung sejak bulan November 2022 sampai bulan Maret 2023.

Tabel 3. 1 Waktu Penelitian

| No | Keterangan | Nov 2022 | | | | Des 2022 | | | | Jan 2023 | | | | Feb 2023 | | | | Mar 2023 | | | |
|----|----------------------------|----------|---|---|---|----------|---|---|---|----------|---|---|---|----------|---|---|---|----------|---|---|---|
| | | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Analisis Masalah | █ | | | | | | | | | | | | | | | | | | | |
| 2 | Studi Literatur | █ | | | | | | | | | | | | | | | | | | | |
| 3 | Pengumpulan Data | █ | | | | █ | | | | | | | | | | | | | | | |
| 4 | Preprocessing | | | | | █ | | | | █ | | | | | | | | | | | |
| 5 | Implementasi Algoritma | | | | | | | | | █ | | | | █ | | | | | | | |
| 6 | Perhitungan Manual (Excel) | | | | | | | | | █ | | | | █ | | | | | | | |
| 7 | Evaluasi | | | | | | | | | | | | | █ | | | | | | | |

3.4 Prosedur Penelitian

Tahapan prosedur penelitian dapat dilihat pada gambar dibawah :



Gambar 3. 1 Prosedur Penelitian

Berdasarkan Gambar 3.1 dapat dilihat bahwa pada penelitian ini mencakup tujuh tahapan yang dilakukan secara berurutan, dimulai dari analisis masalah, studi literatur, pengumpulan data, perhitungan manual menggunakan *excel*, implementasi algoritma pada proses ini pengolahan atau teknik *clustering* menggunakan perbandingan algoritma *K-Means* dan *K-Medoids*, dan terakhir *evaluasi* dari pengujian yang telah dilakukan.

3.5 Analisis Masalah

Setelah melakukan analisis pada Badan Penanggulangan Bencana Daerah (BPBD) belum dapat mengoptimalkan pelayanan terhadap korban bencana, misalnya lamanya datang bantuan karena terbatasnya peralatan dan makanan pada daerah bencana. Sedangkan dengan adanya pemetaan resiko bencana menjadi sangat penting dalam penataan penanggulangan bencana yang terarah dan tepat, untuk kesiapsiagaan pengevaluasian kerusakan tempat tinggal serta meningkatkan kesadaran akan resiko bencana yang tinggi. Maka diperlukannya pengolahan data untuk mengetahui daerah Kabupaten atau Kota di Jawa Barat yang sering terjadi bencana berdasarkan jumlah kerusakannya.

3.6 Studi Literatur

Penelitian ini dimulai dengan melakukan studi literatur yaitu mempelajari jurnal karya ilmiah yang berkaitan dengan teknik data *mining* dengan metode *clustering* menggunakan algoritma *K-Means* dan *K-Medoids* yang bersumber dari internet dan perpustakaan.

3.7 Pengumpulan Data

Pada ditahap ini proses yang dilakukan adalah dengan mencari data pada *website*.opendata.Jabarprov dan data yang didapat merupakan data jumlah kerusakan tempat tinggal akibat bencana dari rentang tahun 2012-2021 yang berjumlah 1620 data.

3.8 Preprocessing

Sebelum melakukan pengolah data terlebih dahulu dilakukan *preprocessing* data. *Preprocessing* data terdiri dari seleksi data, pembersihan data, transformasi data, dan Normalisasi Data.

1. Seleksi Data

Penyeleksian data dilakukan karena data yang didapatkan tidak semuanya digunakan dalam penelitian ini. Seleksi data merupakan proses penyeleksian data dari atribut yang tidak di inginkan, ada beberapa atribut yang tidak relevan atau tidak dibutuhkan. Pada penelitian ini terdapat 3 atribut yang digunakan diantaranya yaitu Nama Kabupaten atau Kota, kondisi kerusakan, dan jumlah kerusakan.

Tabel 3. 2 Seleksi Data

| nama_kabupaten_kota | kondisi_kerusakan | jumlah_kerusakan |
|---------------------|-------------------|------------------|
|---------------------|-------------------|------------------|

| | | |
|-----------------|--------------------|-----|
| KABUPATEN BOGOR | HANCUR | 0 |
| KABUPATEN BOGOR | RUSAK BERAT | 240 |
| KABUPATEN BOGOR | RUSAK SEDANG | 80 |
| ... | ... | ... |
| ... | ... | ... |
| ... | ... | ... |
| KOTA BANJAR | RUSAK RINGAN | 4 |
| KOTA BANJAR | TERANCAM | 0 |
| KOTA BANJAR | TERENDAM/TERTIMBUN | 5 |

2. Transformasi Data

Pada tahap proses transformasi perubahan data menjadi data yang dapat diolah menggunakan algoritma yang akan dipakai, data ditransformasikan ke dalam bentuk yang sudah ditambahkan dari tahun 2012-2021.

Tabel 3. 3 Transformasi Data

| Nama Kabupaten | Hancur | Rusak Berat | Rusak Sedang | Rusak Ringan | Terancam | Terendam |
|--------------------|--------|-------------|--------------|--------------|----------|----------|
| Kabupaten Bogor | 26 | 1168 | 2308 | 5680 | 1431 | 19753 |
| Kabupaten Sukabumi | 57 | 2648 | 3769 | 10007 | 1756 | 2123 |
| Kabupaten Cianjur | 18 | 725 | 438 | 2254 | 3198 | 2504 |
| ... | ... | ... | ... | ... | ... | ... |
| Kota Cimahi | 4 | 86 | 68 | 170 | 18 | 5160 |
| Kota Tasikmalaya | 5 | 106 | 533 | 1228 | 48 | 315 |
| Kota Banjar | 0 | 130 | 616 | 435 | 28 | 24 |

3. Pembersihan Data

Pada tahap pembersihan data di lakukan dengan memisahkan antara data yang digunakan dan data yang tidak diperlukan. Kemudian data yang tidak diperlukan akan dibersihkan agar data menjadi efektif saat melakukan proses perhitungan.

```

Nama Kabupaten    0
Hancur            0
Rusak Berat      0
Rusak Sedang     0
Rusak Ringan     0
Terancam         0
Terendam         0
dtype: int64

```

Gambar 3. 2 Pembersihan Data

4. Normalisasi Data

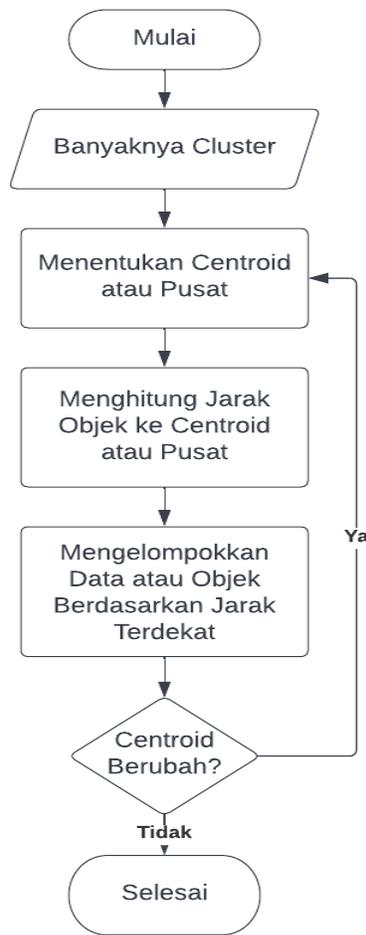
Tahapan *preprocessing* data yang selanjutnya, melakukan normalisasi data, Normalisasi data digunakan untuk mengubah nilai-nilai dari *variable* agar dapat diukur dalam skala umum.

| | Nama Kabupaten | Hancur | Rusak Berat | Rusak Sedang | Rusak Ringan | Terancam | Terendam |
|---|----------------|----------|-------------|--------------|--------------|----------|----------|
| 0 | 3 | 0.093525 | 0.435976 | 0.611333 | 0.565344 | 0.230546 | 0.095297 |
| 1 | 15 | 0.205036 | 1.000000 | 1.000000 | 1.000000 | 0.282906 | 0.010139 |
| 2 | 5 | 0.064748 | 0.267149 | 0.113860 | 0.221195 | 0.515225 | 0.011979 |
| 3 | 0 | 0.226619 | 0.225229 | 0.157223 | 0.242692 | 0.538263 | 1.000000 |
| 4 | 7 | 1.000000 | 0.660823 | 0.715084 | 0.373882 | 0.339133 | 0.035599 |

Gambar 3. 3 Normalisasi Data

3.9 Implementasi Algoritma

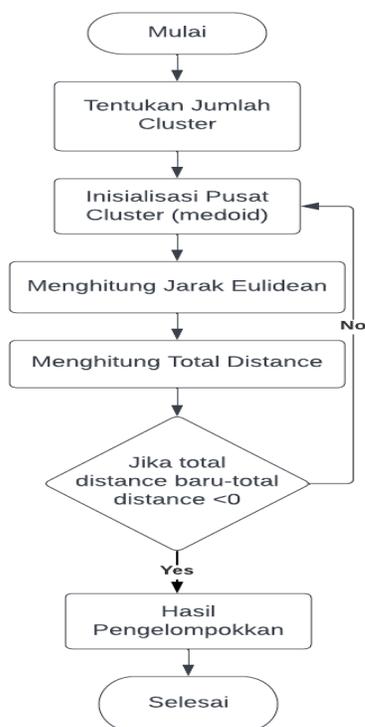
Setelah melakukan tahap *preprocessing* selanjutnya implementasi, melakukan perhitungan menggunakan algoritma *K-Means*, dilanjut menggunakan perhitungan *K-Medoids* dengan bahasa pemrograman python. Berikut tahapan algoritma *K-Means* dapat dilihat pada Gambar 3.4 :



Gambar 3. 4 Flowchart K-Means

Berdasarkan pada Gambar 3.4 dapat dijelaskan langkah awal yang perlu dilakukan yaitu tentukan K data sebagai *centroid*, K adalah jumlah *cluster* yang ditentukan oleh penulis. Lalu pengelompokkan nilai titik tengah, untuk menentukan nilai awal *centroid* dilakukan secara acak. Setelah itu menghitung jarak obyek ke *centroid*. Lalu, melakukan perhitungan pusat klaster yang baru. Pusat klaster yang baru ditentukan berdasarkan pengelompokkan anggota yang ada disetiap klaster. Lalu yang terakhir pengulangan langkah 1-4 sampai centroid tidak mengalami perubahan.

Kemudian untuk tahapan menggunakan algoritma *K-Medoids* yaitu pada gambar 3.5 sebagai berikut :



Gambar 3. 5 Flowchart K-Medoids

Berdasarkan Gambar 3.5 dapat dijelaskan langkah awal yang perlu dilakukan yaitu tentukan nilai K yang akan menjadi cluster, kemudian alokasikan setiap data (objek) ke pusat *cluster* dengan perhitungan persamaan *Euclidean Distance*. Kemudian menghitung total jarak semua data dalam cluster. Kemudian menghitung simpangan (S) untuk menghitung nilai jarak total yang baru – jarak total yang lama. Jika nilai $S < 0$, maka objek dalam klaster akan ditukar dengan data klaster untuk membentuk sebuah kelompok baru yang terdiri dari k objek sebagai medoids. Ulangi langkah 3 sampai 5 sampai tidak ada perubahan pada medoid, sehingga diperoleh sebuah klaster dengan anggota *cluster* yang telah ditentukan.

Pada pengelompokan data kerusakan tempat tinggal akibat bencana sebanyak 1620 data menggunakan perbandingan algoritma *K-Means* dan *K-Medoids* memiliki nilai yang tidak pasti, maka untuk mengetahui kualitas *clustering* dari hasil yang diperoleh, dapat menggunakan perbandingan hasil dari algoritma lain. Oleh karena itu, uji hasil pengolahan merupakan salah satu metode untuk menentukan kualitas *clustering*. Uji pengolahan *clustering* menggunakan metode *silhouette coefficient*. pengujian dilakukan untuk menentukan tingkat akurasi algoritma *K-Means* dan *K-Medoids*.

3.9 Perhitungan Manual (Excel)

Pada proses perhitungan manual penulis menggunakan aplikasi *Microsoft Excel* dan mengimplementasikan algoritma *K-Means* dan *K-Medoids* dalam melakukan *Clustering*. Data akan dikelompokkan menjadi 2 yaitu Tinggi (rawan), dan Rendah (aman).

a. Algoritma *K-Means*

Langkah-langkah dalam algoritma *K-Means Clustering* sebagai berikut :

1. Menentukan jumlah kluster yang akan digunakan.
2. Memilih data secara acak sebanyak jumlah *cluster* sebagai pusat *cluster* (*centroid*)
3. Mengitung jarak setiap data dengan *centroid*, menggunakan rumus jarak *Euclidean distance* :

$$D_{ij} = \sqrt{(X_{1j} - V_{1j})^2 + \dots + (X_{ij} - V_{kj})^2} \quad (3.1)$$

Dimana :

D_{ij} = jarak data ke i ke pusat kluster ke k

X_{ij} = dara ke i , pada atribut ke j

V_{kj} = titik pusat ke k , pada atribut ke j

4. Melakukan perhitung ulang pusat *cluster* (*centroid*) dengan keanggotaan *cluster* yang baru, berikut rumus untuk menghitung *centroid* :

$$C_1 = \frac{1}{M} \sum_{j=1}^M x_j \quad (3.2)$$

Jika pusat *cluster* tidak berubah maka proses klasterisasi selesai, jika pusat kalster masih berubah ulangi langkah ke 4 sampai pusat *cluster* tidak berubah lagi.

b. Algoritma *K-Medoids*

Algoritma *K-Medoids* menggunakan metode partisi *clustering* untuk mengelompokkan sejumlah n objek menjadi sebuah *cluster*. *Cluster* di buat dengan cara menghitung kedekatan yang dimiliki antara medoid dengan objek non medoid. Langkah-langkah dalam perhitungan algoritma *k-medoid* antara lain.

1. Inialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (obek) ke kluster terdekat menggunakan persamaan ukuran jarak *Euclidian Distance* dengan persamaan:
3. (obek) ke kluster terdekat menggunakan rumus jarak *Euclidian Distance* dengan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.3)$$

Dimana:

$d(x,y)$ = jarak antara data ke-I da data ke j

x_{i1} = nilai atribut ke satu dari data ke-i

y_{j1} = nilai atribut ke satu dari ke-j

n = jumlah atribut yang digunakan.

4. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru.
5. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoid baru.
6. Hitung total simpangan (S) dengan menghitung nilai total *distance* baru total *distance* lama. Jika $S < 0$, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai medoid.
7. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan cluster beserta anggota cluster masing-masing.

3.10 Evaluasi

Metode *Silhouette Coefficient*. Metode ini merupakan metode evaluasi *cluster* untuk melihat kualitas objek dalam suatu *cluster*. Berikut langkah – langkah perhitungannya menurut (Nurlaela et al., 2020) diantaranya:

- 1) Menghitung jarak rata – rata dari suatu objek ke- i dengan semua objek lain yang berada dalam satu *cluster*.

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3.4)$$

- 2) Menghitung rata – rata dari objek ke- i tersebut dengan semua objek pada *cluster* yang lainnya, dan ambil nilai terkecilnya.

$$d(i) = \frac{1}{[A]} \sum_{j \in C} d(i, j) \quad (3.5)$$

- 3) Nilai *Silhouette Coefficient* nya adalah jumlah $s(i)$ diperoleh dengan menggunakan $a(i)$ dan $b(i)$ sehingga dapat dirumuskan:

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (3.6)$$

Dimana:

$s(i)$ = Nilai *Silhouette*.

$a(i)$ = Rata – rata jarak antara i dengan semua objek pada cluster.

$b(i)$ = Rata – rata jarak antara data i pada semua objek pada cluster.

$d(i, j)$ = Jarak antara objek i dengan j .

