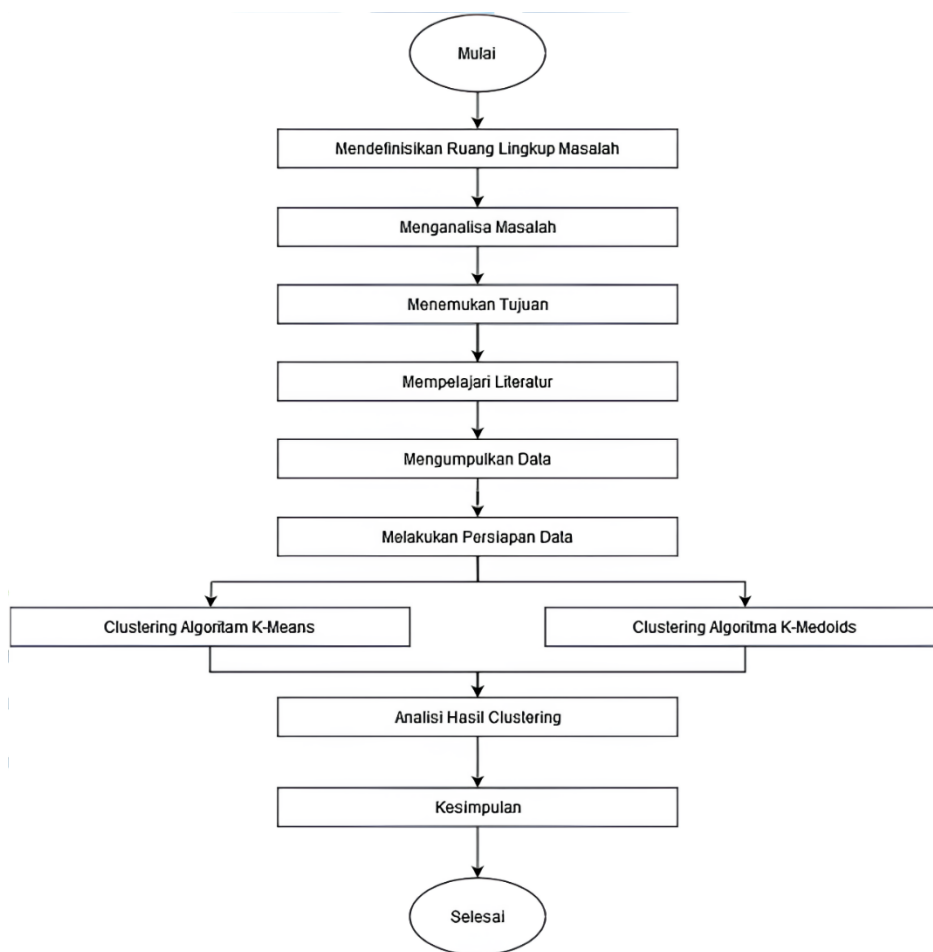


# BAB III

## Metodologi Penelitian

### 3.1 Alur Rancangan Penelitian

Tahapan pengerjaan penelitian merupakan langkah-langkah yang akan dilakukan peneliti dalam melakukan penelitian. Skema alur tahapan yang digunakan penelitian ini ditunjukkan gambar 1.



Gambar 1. Alur Penelitian

Gambar 1. Adalah proses penelitian yang melibatkan peninjauan serta evaluasi kritis terhadap sumber-sumber literatur yang telah ada. Tujuan utama dari studi literatur adalah untuk memahami dan menggambarkan keadaan penelitian terkini di bidang yang berkaitan dengan topik yang diteliti, dengan melakukan literatur review, peneliti dapat menemukan pengetahuan terkini, mengidentifikasi kesenjangan penelitian yang ada, serta menyusun dasar teoretis yang kuat untuk penelitian yang akan dilakukan.

### **3.2 Tahapan awal Alur Rancangan Penelitian adalah:**

#### 1. Start

Tahapan awal sebagai landasan penting saat memulai proses penelitian.

#### 2. Mendefinisikan Ruang Lingkup Masalah

Tahap mendefinisikan ruang lingkup untuk mendefinisikan rumusan dan batasan masalah yang akan diteliti, ruang lingkup penelitian ini akan berfokus penggunaan algoritma clustering K-Means dan K-Medoids serta melakukan perbandingan antara kedua algoritma. Perbandingan dilakukan dengan membandingkan akurasi clustering yang telah terbentuk menggunakan metoda Davies Bouldin-Index.

#### 3. Menganalisa Masalah

Tahap menganalisa masalah dilakukan pemahaman masalah yang telah ditentukan ruang lingkup atau batasannya, didapatkan analisa masalah yang akan dibahas di penelitian ini adalah melakukan clustering data Covid-19 di

Indonesia. Data ini dapat dimanfaatkan untuk mendapatkan informasi terkait kasus penyebaran Covid dengan melakukan clustering. Kemudian dilakukan juga perbandingan antara dua algoritma mana yang paling tepat digunakan dalam melakukan clustering data Covid-19.

#### 4. Menemukan Tujuan

Tahap menemukan tujuan ini berdasarkan pemahaman dari masalah yang dibahas, maka ditentukan tujuan yang akan dicapai dari penelitian ini adalah melakukan perbandingan algoritma clustering K-Means dan K-Medoids pada data Covid-19 di Indonesia, mendapatkan hasil akhir clustering data Covid-19 beserta perbandingan kinerja algoritma clustering dengan metode Davies Bouldin-Index, kemudian menentukan algoritma yang paling tepat untuk digunakan clustering.

#### 5. Mempelajari Literatur

Tahap mempelajari literatur, dilakukan pembelajaran dari literatur yang telah ada untuk dapat dijadikan dasar atau rujukan dalam penelitian ini. Di penelitian ini menggunakan beberapa studi literatur yang didapat berupa bahan yang akan diterbitkan contohnya jurnal, buku dan literatur lainnya guna mendukung dasar dan sebagai rujukan dalam penelitian ini.

#### 6. Mengumpulkan Data

Tahap pengumpulan data dilakukan pengamatan melalui website milik pemerintah yaitu [www.covid19.go.id](http://www.covid19.go.id). Data yang digunakan untuk clustering adalah data kasus penyebaran Covid-19 di Indonesia yang meliputi seluruh

wilayah Indonesia dengan variabel Positif, Sembuh dan Meninggal. Data tersebut dimulai dari awal kemunculan Covid-19 di Indonesia hingga tahun 2022.

#### 7. Melakukan Persiapan Data

Tahap persiapan data, data dibersihkan, dipilih, dan diubah. Tahap pembersihan data awal dilakukan untuk memastikan bahwa semua dataset telah lengkap dan siap untuk diolah sesuai kebutuhan. Tujuan dari tahap ini adalah untuk memastikan bahwa tidak ada kolom bernilai yang kosong atau tidak lengkap.

#### 8. Clustering Algoritma

Clustering juga dapat didefinisikan sebagai identifikasi kelas objek yang memiliki kemiripan. Clustering adalah pengelompokan terhadap daftar, yang berfungsi untuk memperhatikan dan membentuk kelas objek yang memiliki kemiripan. Dengan menggunakan metode clustering kita dapat lebih lanjut menemukan kepadatan dan jarak daerah dalam objek ruang serta menemukan pola distribusi dan korelasi antara atribut. Penggunaan teknik clustering menghasilkan penentuan beberapa cluster catatan data yang memiliki karakteristik yang mirip. Nilai kesamaan antara dua objek yang sebanding dapat dihitung dengan berbagai cara. Salah satunya adalah jarak geometris, yang menghitung jarak dua poin dengan mengetahui nilai masing-masing atribut mereka. Untuk menentukan apakah dua vektor fitur yang dinyatakan

dengan ranking memiliki kesamaan atau ketidaksamaan, metode yang biasa digunakan adalah jarak.

### **K-Means Clustering**

Algoritma K-Means merupakan metode pengclusteran secara partitioning yang memisahkan data ke dalam kelompok yang berbeda. Dengan partitioning secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke clusternya. Penggunaan algoritma K-Means dalam melakukan proses clustering tergantung dari data yang ada dan konklusi yang ingin di capai.(Rahayu, Prianto, and Novia 2021).

K-means mencoba meminimalkan perbedaan kesalahan kuadrat antara rata-rata cluster dan titik data di cluster itu. Misalkan kita punya beberapa titik data n-dimensi yang ingin dikelompokkan pengguna dalam k jumlah cluster dengan  $\mu_k$  sebagai mean dari cluster tersebut, maka k-means direpresentasikan sebagai:

$$\left[ \sum_{k=1}^k \sum_{x \in c_k} \|x_i - u_k\|^2 \right]$$

Dimana  $x_i$  adalah himpunan titik data dengan  $i = 1, 2, 3, \dots, n$ , menjadi dikelompokkan dalam sebuah:

- 1) cluster dari satu set cluster yang diberikan sebagai  $c_k$  dengan  $k = 1, 2, 3, \dots, k$ . Untuk mengurangi kesalahan kuadrat, k-means mengalokasikan

pola ke kluster k yang awalnya dipartisi. Dalam hal keanggotaan cluster yang tidak menentukan, k-means terus ulangi langkah-langkah berikut :Tetapkan setiap pola ke cluster terdekat dan buat partisi baru

2) Hitung rata-rata cluster baru.

### **K-Medoids Clustering**

K-Medoids merupakan suatu algoritma yang digunakan untuk menemukan medoids didalam sebuah kelompok (cluster) yang merupakan titik pusat dari suatu kelompok (cluster). Algoritma K-Medoids lebih baik dibandingkan dengan K-Means karena K-Medoids kita menemukan k sebagai objek yang representatif untuk meminimalkan jumlah ketidaksamaan objek data, sedangkan K-Means menggunakan jumlah jarak euclidean distances untuk objek data.(Sindi et al. 2020). Langkah-Langkah Penyelesaian K-Medoids adalah sebagai berikut:

- 1) Inisialisasikan pusat cluster sebanyak jumlah cluster (k).
- 2) Setiap data atau objek dialokasikan ke cluster terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan rumus persamaan.
- 3) Pilih objek masing-masing cluster secara acak sebagai kandidat medoid baru.

- 4) Hitung jarak setiap objek yang terdapat masing-masing cluster dengan calon medoid baru.
- 5) Hitung total simpangan (S) dengan menghitung nilai total jarak baru – total jarak lama. Jika didapatkan  $S < 0$ , tukarlah objek dengan data cluster untuk membuat sekumpulan k objek baru sebagai medoid.
- 6) Ulangi langkah 3 sampai dengan 5 hingga tidak terjadi perubahan medoid, sehingga diperoleh cluster.

#### **Davies-Bouldin Index (DBI)**

Davies-Bouldin Index (DBI) ialah teknik suatu ukuran yang akan digunakan untuk menentukan jumlah cluster yang terbaik setelah proses pengklasteran terselesaikan. Metode DBI ini bermaksud untuk mengoptimalkan perbedaan antara klaster yang satu dengan klaster yang lain serta mencoba meminimalkan jarak antar objek suatu klaster, dan juga dalam metode DBI ini bisa menjadi parameter setiap algoritma clustering untuk membandingkan proses cluster algoritma mana yang lebih optimal. Dalam metode ini semakin kecil nilai yang mendekati 0, semakin optimal clustering yang dihasilkan. (Kaligis and Yulianto 2022)

#### **9. Analisis Hasil Clustering**

Langkah analisis hasil kelompok ini, hasil kelompok yang dibentuk dievaluasi untuk mengetahui penyebaran Covid-19. Selain itu, hasil

kelompok dievaluasi dengan metode Davies Bouldin Index, yang merupakan ukuran yang menilai kinerja kelompokan. Digunakan sebagai metrik pengelompokan, DBI menunjukkan korelasi positif untuk kasus "dalam kelas" dan korelasi negatif untuk kasus "antara kelas". Validasi eksternal dan internal adalah dua kategori utama validasi yang digunakan untuk mengevaluasi kinerja hasil kelompok.

#### 10. Kesimpulan

Tahap kesimpulan ini, hasil clustering kedua algoritma dibandingkan dan ditarik kesimpulan akhir. Ini menghasilkan kesimpulan bahwa algoritma mana yang paling sesuai untuk menggabungkan data penyebaran COVID-19 di Indonesia dengan variabel yang positif, sembuh, dan meninggal.