

# BABI

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Covid-19 ialah epidemi mematikan dikarenakan oleh virus *Corona* yang merupakan penyakit endemi di seluruh dunia, pertama kali di temukan Di Wuhan oleh seorang Dokter yang bernama Zhang Jixian adalah orang pertama yang melaporkan dan menyadari bahwa sifat virus sangat menular (Bavel et al., 2020). Penyebaran virus *Corona* di Indonesia pertama kali ditemukan oleh pemerintah pada 2 Maret 2020 (Dwitri et al., 2020), Ketika penyebaran virus corona masih merajalela, pemerintah Indonesia mengambil langkah untuk mendorong masyarakat agar patuh terhadap pembatasan fisik atau menerapkan prinsip jarak terkendali guna mengurangi penyebaran virus tersebut (Rizal et al., 2021). dan prevalensi kasus COVID-19 meningkat setiap hari, dengan dampak negatif yang signifikan di semua wilayah. Aceh merupakan wilayah yang terkena dampak COVID-19 hingga akhir tahun 2022 berdasarkan data dari covid19.acehprov.go.id yang terkonfirmasi berjumlah 44.853 orang, sembuh 42356 orang dan meninggal 2265 orang. Virus *Corona* sangat lah berbahaya dan memiliki tingkat penyebaran yang tinggi, namun masyarakat kurang mengetahui daerah yang terkonfirmasi memiliki pasien yang terjangkit virus ini.

Terkait hal tersebut, penulis ingin melakukan penelitian untuk mendapatkan klasterisasi wilayah di provinsi Aceh yang dapat memetakan zona resiko penyebaran COVID-19. Untuk klasterisasi, ada beberapa algoritma yang dapat digunakan namun pada penelitian ini penulis akan menggunakan algoritma *K-Means* dan C4.5. Algoritma *K-Means* ialah teknik klasifikasi data *non-hierarchical* yang berguna untuk membagi data menjadi satu atau beberapa klaster atau grup (Harahap et al., 2022). Prinsipnya adalah mengelompokkan data dengan ciri-ciri serupa ke dalam klaster yang sama, sementara data yang memiliki ciri-ciri yang berbeda dikelompokkan ke dalam klaster yang berbeda, sedangkan algoritma C4.5

adalah algoritma yang dipakai untuk membentuk pohon keputusan dengan melakukan klasifikasi pada dataset (Faid et al., 2023). Adapun kelebihan dari algoritma *K-Means* adalah mempermudah pengimplementasian, relatif cepat dan fleksibel serta menggunakan prinsip sederhana. Selain itu algoritma *K-Means* mempunyai kekurangan seperti perlu menentukan nilai *k* secara manual, sangat bergantung pada inisialisasi awal dan dapat terjadi *curse of dimensionality*. Adapun kelebihan dari algoritma C4.5 ialah memiliki tingkat akurasi yang dapat diterima, efisien, populer dan juga mudah untuk diinterpretasikan selain itu algoritma C4.5 juga memiliki kekurangan yaitu hanya dapat digunakan untuk sampel yang dapat disimpan secara menyeluruh pada waktu yang sama di dalam memori.

Penelitian terdahulu pernah dilakukan oleh (Hamdani et al., 2020) dengan judul “Implementasi Metode *K-Means Clustering* Dalam Pengelompokan Penyebaran COVID-19 Di Surabaya”, dari penelitian yang telah mereka lakukan ialah dengan menggunakan dengan metode *K-Means* dan menggunakan *Silhouette Index* untuk mendapatkan hasil yang akurat dari pengklasifikasian yang mereka lakukan secara keseluruhan menghasilkan struktur yang baik tetapi pada uji coba 2 *cluster* menghasilkan struktur yang kuat dengan *Silhouette Index* sebesar 0.8021. Pada penelitian ini, pengelompokan data *cluster* daerah di Aceh berdasarkan data positif, sembuh, dan meninggal menggunakan algoritma *K-Means*. Dalam penelitian ini metode algoritma C4.5 dipakai untuk menampilkan aturan dalam bentuk pohon keputusan dan menggunakan empat label pengelompokan yang digunakan yaitu *cluster* darurat (area hitam), *cluster* tinggi (area merah), *cluster* sedang (area kuning), dan *cluster* rendah (area hijau). Maka berdasarkan penjelasan di atas, dalam pengajuan tugas akhir penulis membahas tentang “**Penerapan Algoritma *K-Means* dan C4.5 Dalam Pengelompokan COVID-19 Di Aceh**”.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang di atas, maka yang menjadi rumusan masalah dalam penelitian ini adalah :

1. Bagaimana cara membuat sistem pengelompokan menggunakan algoritma *K-means* dan C 4.5 pada kasus COVID-19 di Aceh?

2. Bagaimana hasil uji dari jumlah kluster dalam pengimplementasian algoritma *K-means* dan algoritma C 4.5?

### 1.3 Batasan Masalah

Penelitian ini ditentukan pada ruang lingkup tertentu antara lain :

1. Data yang diambil dari tanggal 2 Maret 2020 sampai dengan tanggal 9 September 2023
2. Hanya kasus COVID-19 dari 23 kabupaten di provinsi Aceh.
3. Metode yang digunakan *K-Means* dan C4.5
4. Hasil yang didapatkan hanya pengelompokan *cluster* yaitu tinggi, sedang dan rendah dan area COVID -19 yaitu label zona hijau , zona kuning , zona merah dan zona hitam.
5. Data diambil hanya data yang positif (terkonfirmasi), sembuh, dan meninggal.

### 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang diuraikan diatas, maka tujuan penelitian ini adalah sebagai berikut :

1. Untuk mengklasifikasikan *cluster* dan area COVID-19 di provinsi Aceh menggunakan metode *K-Means* untuk mendapatkan 3 *cluster* yang diinginkan yaitu “tinggi, sedang dan rendah” dan juga untuk mengetahui bentuk pohon (*tree*) menggunakan C4.5 yang menghasilkan 4 label yaitu : zona hitam (darurat), zona merah (tinggi), zona kuning (sedang), dan zona hijau (aman).
2. Untuk mengetahui data uji dari *cluster* yang valid dan sesuai dengan pengklasifikasian yang telah dilakukan.

### 1.5 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini antara lain :

1. Memperoleh pengalaman menarik, wawasan, juga ilmu pengetahuan yang baru bagi penulis dan pembaca tentang pengelompokan memakai algoritma *K-Means* dan algoritma C4.5.
2. Dapat memberi informasi bagi masyarakat serta pemerintah Aceh dalam membuat kebijakan yang seharusnya.

3. Dapat membantu masyarakat untuk menjauhi daerah yang rawan terpapar COVID-19

