

BAB I

PENDAHULUAN

1.1 Latar Belakang

Virus Corona (covid-19) merupakan penyakit menular yang dapat menyerang manusia dan hewan. Virus ini ditemukan sekitar akhir Desember 2019 di Provinsi Wuhan, China. Saat ini, seluruh dunia sedang memerangi, menghentikan, dan mengalahkan penyebaran virus corona. Berdasarkan statistik World Health Organization hingga 27 Oktober 2021, sebanyak 4.241.849 kasus covid-19 yang ada di Indonesia dengan 143.249 kasus kematian [1]. Akibat pandemi covid-19 ini mempengaruhi ekonomi dunia tidak hanya Indonesia saja.

Angka kasus positif covid-19 yang terus meningkat di setiap kabupaten atau kota di Pulau Jawa karena penduduk terbanyak di Indonesia berada di pulau Jawa. Beberapa kabupaten/kota di Pulau Jawa dengan skala kasus positif tertinggi akan membutuhkan perhatian khusus. Terdapat beberapa kabupaten/kota dengan prevalensi penularan virus corona yang tinggi, yang seharusnya diberikan perlakuan khusus agar tingkat penyebarannya tidak semakin parah dan menyulitkan pemerintah Indonesia untuk mendefinisikan contoh penanganan virus corona [2].

Langkah pemerintah untuk membatasi penyebaran covid-19 antara lain penetapan Program Pembatasan Kegiatan Masyarakat (PPKM). Jika jumlah kasus positif bertambah, PPKM akan dilaksanakan dengan menggunakan data variabel yang dikumpulkan dari kasus positif, dosis vaksin pertama, dan dosis vaksin kedua. Selain itu, lonjakan kasus baru ini disebabkan oleh masih banyaknya penduduk yang menolak menggunakan masker saat bepergian. Penggunaan masker secara langsung dapat menghambat penularan infeksi karena masker memenuhi peraturan kesehatan [3].

Data yang disajikan oleh *covid19.go.id* hanya agregat data per-provinsi, namun perlu informasi sebaran kasus covid perkabupaten/kota. Penelitian ini bertujuan untuk mengelompokkan data sebaran covid-19 pada kabupaten/kota di pulau Jawa sehingga dapat membantu pembentukan klaster zona berdasarkan karakteristik positif, vaksin dosis pertama, dan vaksin dosis kedua [4]. *Clustering* adalah teknik untuk mengkategorikan data berdasarkan kesamaan karakteristik di seluruh kelompok data. Metode *K-Means*, metode *K-Means++*, pendekatan *K-Modes*, *Hierarchical clustering*, dan teknik lainnya adalah beberapa cara pengelompokan data. *K-Modes* lebih berguna untuk mengolah data kategorikal.

Penggunaan *K-Means* untuk klasterisasi data covid-19 merupakan salah satu pendekatan analisis cluster *non-hierarchical* yaitu strategi *cluster* dimana jumlah *cluster* diatur secara manual. Tujuan dari *cluster* ini adalah untuk membagi objek yang diamati menjadi satu atau lebih kelompok tergantung pada karakteristiknya. Objek dengan karakteristik yang sama akan ditempatkan dalam *cluster* yang sama, sedangkan objek dengan karakteristik yang berbeda akan dicampur dengan *cluster* lainnya [5]. Dalam kajian yang menganalisis *K-Means*, Siti Azizatus Sholihah mengelompokkan provinsi berdasarkan prevalensi covid-19 di setiap provinsi di Indonesia. Kesimpulan dari penelitian ini adalah rata-rata *Silhouette Coefficient* (SC) untuk $k = 2$ adalah 0,74 dengan 3 data tidak akurat karena 3 negatif. Namun *Silhouette Coefficient* = 0,74 termasuk dalam struktur kokoh. Provinsi Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Banten, dan Riau tergolong rawan penyebaran virus covid-19 berdasarkan teknik *K-Means*. Sementara 28 provinsi dinyatakan bebas dari risiko penularan covid-19. Kekurangan pada metode *K-Means* yaitu pada pemilihan centroid awal. Metode *K-Means++* memilih centroid awal yang dengan cara yang berbeda dengan *K-Means* [6].

Beberapa latar belakang diatas maka perlunya untuk informasi sebaran mengenai kasus covid-19 pada setiap kabupaten/kota di pulau Jawa dan perlu menguji *cluster* dari metode *K-Means* dan *K-Means++* untuk mencari *cluster* yang optimal dari kedua metode tersebut. Penelitian ini akan menggunakan variabel positif, vaksin dosis pertama, dan vaksin dosis kedua untuk mengetahui perkembangan kasus covid-19 dengan mengelompokkan zona penyebaran virus corona, yang dapat memberikan informasi status zona berdasarkan hasil *cluster* yang digunakan. Hasil dari *cluster* tersebut akan dievaluasi dengan menggunakan metode *silhouette coefficient* untuk menentukan cluster mana yang memiliki kualitas lebih baik.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka pokok – pokok permasalahan dapat dirumuskan sebagai berikut:

1. Bagaimana *cluster* tingkat penyebaran covid-19 setiap kabupaten dan kota di Pulau Jawa menggunakan data pasien covid-19 ?
2. Bagaimana hasil kualitas *cluster* dari metode *K-Means* dan *K- Means++* menggunakan data kasus covid-19 di Pulau Jawa ?

1.3 Batasan Masalah

Berdasarkan latar belakang, maka pada penelitian ini memiliki batasan masalah dari penelitian ini yaitu :

1. Menggunakan data pasien covid-19 terbaru pada tanggal 25 Maret 2021 - 09 Febuari 2022.
2. Menggunakan data setiap provinsi di Indonesia dengan variabel positif, vaksin dosis pertama, dan vaksin dosis kedua.
3. Nilai k yang akan diuji yaitu k = 3 sampai k = 6.
4. Menggunakan *Silhouette Coefficient* untuk menguji setiap *cluster*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini yaitu :

1. Untuk mengetahui *cluster* penyebaran covid-19 kabupaten dan kota yang rawan terpapar kasus covid-19 menggunakan metode *K-Means* dan *K-Means++*.
2. Untuk mengetahui hasil kualitas *cluster* dari metode *K-Means* dan *K-Means++* akan diuji menggunakan *Silhouette Coefficient*.

1.5 Manfaat Penelitian

1. Manfaat untuk bidang kajian *clustering*
Mendapatkan informasi *cluster* ideal antara metode *K-Means* dan *K-Means++* dari hasil pengujian menggunakan *Silhouette coefficient*.
2. Manfaat untuk masyarakat terkait informasi covid-19
Dapat mengetahui tingkat penyebaran kasus covid-19 kabupaten dan kota di Pulau Jawa.