

BAB I

PENDAHULUAN

Bab ini berisikan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan skripsi.

1.1 Latar Belakang

Dalam kehidupan sehari-hari, pengelompokan suatu objek sering dijumpai, baik berupa benda, benda mati maupun benda hidup, atau hal lainnya. Pengelompokan didasarkan pada suatu kesamaan. Misalkan di lingkungan universitas, fakultas merupakan contoh kelompok yang berisi program-program studi yang memiliki kesamaan dalam hal ilmu yang dipelajari. Atau sebagai contoh lainnya, laki-laki dan wanita yang merupakan contoh pengelompokan manusia berdasarkan kesamaan jenis kelamin.

Dalam ilmu statistika, dipelajari sebuah analisis bernama analisis multivariat. Analisis multivariat merupakan analisis data dengan menggunakan banyak variabel. Analisis multivariat dapat digunakan untuk memecahkan masalah dalam kehidupan sehari-hari, termasuk mengelompokkan objek-objek yang memiliki karakteristik serupa. Analisis yang dapat digunakan adalah analisis *cluster*. Analisis *cluster* didasarkan pada prinsip homogenitas (persamaan) yang tinggi antar anggota dalam satu *cluster* (*within-cluster*) dan heterogenitas (perbedaan) yang tinggi antar *cluster* yang satu dengan lainnya (*between cluster*), tujuannya adalah untuk mengelompokkan objek berdasarkan kemiripannya. Manfaat analisis *cluster* yang diterapkan dalam berbagai bidang, diantaranya meringkas atau mereduksi data, segmentasi pasar konsumen, memahami perilaku pembeli, dll.

Analisis *cluster* yang sering dijumpai biasanya menggunakan struktur data *cross-section*. Data *cross-section* adalah data yang dikumpulkan hanya pada titik waktu tertentu saja untuk mengetahui keadaan pada waktu itu, dengan kata lain data *cross-section* tidak dipengaruhi oleh waktu. Terdapat tipe data jenis lain yang

secara signifikan berbeda dengan *cross-section*, yakni data dipengaruhi oleh waktu yang disebut data deret waktu.

Analisis *cluster* deret waktu merupakan jenis khusus dari analisis *cluster* di mana pengelompokan diterapkan pada data deret waktu. Terdapat beberapa penelitian terdahulu mengenai analisis *cluster* deret waktu diantaranya Aghabozorgi *et al* meninjau penggunaan analisis *cluster* deret waktu pada beberapa waktu yang lalu[1], O. Kobylin and V. Lyashenko melakukan pengelompokan deret waktu menggunakan algoritma K-Means[2], J. Ahn dan J. H. Lee menerapkan algoritma DBSCAN dan K-Means untuk pengelompokan deret waktu[3], A. D. Munthe melakukan pengelompokan provinsi di Indonesia berdasarkan nilai produksi padi dengan analisis *cluster* deret waktu[4], L. Putu Widya Adnyani dan P. Robinson Sihombing melakukan analisis *cluster* deret waktu untuk mengelompokkan provinsi di Indonesia berdasarkan nilai PDRB[5], N. Priatna menganalisis perbandingan algoritma *cluseing* K-Means dan DBSCAN[6], N. James dan M. Menzies melakukan analisis *cluster* pada data deret waktu COVID-19 untuk mengetahui kebijakan negara mana saja yang paling efektif dan tidak efektif dalam mengontrol penyebaran COVID-19[7].

Di awal tahun 2021, dunia terjadi pandemi COVID-19. Pandemi COVID-19 adalah penyebaran global *Coronavirus Disease 2019* (COVID-19) di semua negara. *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) adalah penyebab penyakit menular COVID-19. Tanda dan gejala umum infeksi COVID-19 meliputi gejala gangguan pernapasan akut seperti demam, batuk, dan sesak napas. COVID-19 dimulai pada akhir Desember 2019 dengan ditemukannya kasus pneumonia misterius pertama di Wuhan, Provinsi Hubei, China. Studi epidemiologis menunjukkan bahwa kasus ini terkait dengan pasar ikan dan kerang di Wuhan. Sampai saat ini kasus COVID-19 terus meningkat dan tersebar luas di berbagai wilayah, termasuk Indonesia. Salah satu provinsi dengan kasus aktif COVID-19 yang tinggi yaitu Jawa Barat. Sejauh ini pemerintah terus mengeluarkan kebijakan-kebijakan sebagai upaya mengendalikan laju COVID-19 seperti diantaranya gencar memberlakukan PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat).

Dengan latar belakang di atas, penulis tertarik untuk menerapkan analisis *cluster* data deret waktu untuk mengelompokkan kabupaten/kota di Provinsi Jawa Barat berdasarkan kasus aktif terkonfirmasi COVID-19. Pada skripsi ini akan menggunakan 3 metode pengelompokan yaitu K-Means, Ward, dan *DBSCAN* dengan pengukuran kesamaan yaitu jarak *Dynamic Time Warping*. Tujuan menggunakan tiga metode pada skripsi ini adalah untuk mendapatkan hasil pengelompokan terbaik.

1.2 Rumusan Masalah

Rumusan masalah yang akan dikaji dalam penelitian ini adalah:

1. Bagaimana algoritma menentukan pengelompokan data deret waktu berdasarkan metode K-Means, Ward, DBSCAN?
2. Bagaimana perbandingan hasil akhir pengelompokan Kabupaten/Kota data deret waktu COVID-19 berdasarkan metode K-Means, Ward, dan DBSCAN?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah:

1. Jenis analisis *cluster* deret waktu yang digunakan *whole time series clustering* dengan pendekatan secara langsung dari data mentah.
2. Pengukuran kesamaan menggunakan jarak *Dynamic Time Warping*.
3. Validasi *cluster* menggunakan kriteria siluet.
4. Data yang digunakan dalam penelitian ini yaitu data kasus terkonfirmasi COVID-19 di Provinsi Jawa Barat dari Agustus 2020 sampai Juni 2021.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Menentukan algoritma pengelompokan data deret waktu berdasarkan metode K-Means, Ward, DBSCAN
2. Membandingkan hasil akhir pengelompokan data deret waktu pada kasus terkonfirmasi COVID-19 berdasarkan metode K-Means, Ward dan DBSCAN

1.5 Metode Penelitian

Metode pada penelitian ini adalah:

1. Melakukan studi literatur, yaitu proses di mana penulis akan mengkaji metode K-Means, Ward dan DBSCAN dalam analisis *cluster* data deret waktu. Dengan sumber pustaka berupa buku, jurnal dan skripsi terkait.
2. Penerapan analisis *cluster* pada data kasus terkonfirmasi COVID-19 di Provinsi Jawa Barat dengan bantuan *software R* dan membuat kesimpulan dari hasil yang didapat.

1.6 Sistematika Penulisan

BAB I : PENDAHULUAN

Bab ini memuat latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Bab ini memuat teori-teori yang melandasi pembahasan dalam penulisan skripsi. Secara umum, bab ini mencakup semua yang berkaitan dengan data deret waktu, analisis multivariat, analisis *cluster*, metode K-Means, metode Ward, metode DBSCAN, pandemi COVID-19, dan *software R*.

BAB III : PENGELOMPOKAN DATA DERET WAKTU DENGAN METODE K-MEANS, WARD, DAN DBSCAN

Bab ini memuat pembahasan utama dari pengkajian skripsi, meliputi pembahasan mengenai analisis *cluster* data deret waktu, jarak *Dynamic Time Warping*, serta pengelompokan data dengan metode K-Means, metode Ward, dan metode DBSCAN

BAB IV : STUDI KASUS

Bab ini memuat kasus yang diteliti dan pembahasannya serta perhitungan dengan bantuan *software R* untuk mencari hasil

pengelompokan data deret waktu dengan ketiga metode dan membandingkan ketiganya untuk memperoleh hasil pengelompokan terbaik.

BAB V : PENUTUP

Bab ini berisi kesimpulan dari seluruh pembahasan yang telah dikaji sesuai dengan tujuan yang ingin dicapai. Ada juga diberikan saran untuk pengembangan topik diskusi lebih lanjut.

