

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

World Health Organization (WHO) memberikan informasi mengenai virus baru yang muncul di Wuhan, Cina pada akhir tahun 2019 yaitu *Coronavirus Disease* atau Covid-19. Awal muncul Covid-19 dikarenakan sebagian besar masyarakat Wuhan mengalami gejala *pneumonia* [1]. Proses penyebaran Covid-19 melalui batuk dan bersin melalui perantara yaitu tangan, baju atau lainnya [2].

Penyebaran Covid-19 begitu cepat menyebarkan kewilayah dunia termasuk Indonesia. Di Indonesia khususnya wilayah DKI Jakarta merupakan daerah penyebaran Covid-19 tertinggi. Dari laman resmi Corona Jakarta per tanggal 31 Oktober 2022 tercatat total positif 1.447.780 pasien [3]. Dari banyaknya jumlah pasien yang terkonfirmasi Covid-19 di DKI Jakarta dapat menyebabkan penyebaran yang semakin cepat diseluruh wilayah kecamatan dan kelurahan DKI Jakarta sehingga diperlukan pengelompokan wilayah di DKI Jakarta terhadap penyebaran Covid-19 untuk dapat diketahui titik-titik pusat penyebaran kasus Covid-19 [4]. Dengan mengetahui pusat penyebaran kasus Covid-19 memberikan manfaat bagi pemerintah DKI Jakarta dalam mengelompokkan daerah kedalam kluster sesuai dampak Covid-19 yang dialami daerah tersebut, serta memberikan bantuan secara tepat dari hasil kluster penyebaran Covid-19 di DKI Jakarta. Dalam mengetahui pusat penyebaran Covid-19 di DKI Jakarta dilakukan dengan klusterisasi menggunakan metode *k-means* dan *fuzzy c-means*. Dengan menggunakan klustering yang berarti suatu metode pengelompokan berdasarkan ukuran kedekatan atau kemiripan [4].

Penggunaan metode *K-means* dikarenakan salah satu metode clustering non hirarki yang memiliki waktu komputasi yang relatif cepat [5]. Sedangkan untuk *Fuzzy C-Means* dipilih dikarenakan merupakan teknik pengelompokan yang cocok untuk kemampuan dalam pengenalan pola, data yang tidak lengkap, informasi campuran,

dan dapat memberikan solusi perkiraan lebih cepat [6]. Kedua metode tersebut merupakan kategori *data mining* yang berjenis *unsupervised learning* dengan metode *partitional* klustering. Pada metode *partitional* klustering memiliki titik pusat kluster (*centroid*) pada masing-masing kluster dan memiliki tujuan untuk meminimalkan jarak (*dissimilarity*) dari seluruh data ke masing-masing pusat kluster [7]. Penelitian sebelumnya pernah dilakukan oleh [8] melakukan komparasi metode *k-means* dan *k-medoids* untuk klusterisasi kasus Covid-19 di DKI Jakarta. Dari penelitian tersebut memperoleh hasil kluster optimal dengan jumlah 2 kluster yang terdiri dari kluster 1 memiliki 2 wilayah dan berkategori tinggi. Hasil perbandingan metode tersebut dilakukan evaluasi nilai *Silhouette Coefficient* pada masing-masing metode, sehingga diperoleh nilai 0,207 pada metode *k-means* dan 0,347 pada *k-medoids*.

Pada penelitian ini menggunakan variabel data dependen dan independen. Variabel independen merupakan variabel data-data yang digunakan untuk membuat klusterisasi yang direncanakan. Variabel ini akan memberikan pengaruh terhadap perubahan pada variabel dependen. Variabel independent dalam penelitian ini yang terdiri dari: Positif, sembuh, meninggal, dirawat, dan isoman. Variabel dependen merupakan luaran klusterisasi yang dihasilkan dari model data mining yang dibuat, variabel ini dipengaruhi oleh variabel independent. Variabel dependen yang digunakan dalam penelitian ini yang terdiri dari: kluster merah untuk mengindikasikan daerah yang rawan dengan jumlah pasien Covid-19 terbanyak. Kluster kuning untuk mengindikasikan daerah yang cukup rawan dengan jumlah pasien Covid-19 hampir banyak. Kluster hijau untuk mengindikasikan daerah yang aman dengan jumlah pasien Covid-19 tidak terlalu banyak.

Tujuan melakukan penelitian ini ialah melakukan pengklusteran penyebaran virus Covid-19 di DKI Jakarta dengan menerapkan teknik data mining menggunakan alat bantu bahasa pemrograman *Python* dengan tools *Jupyter Notebook*. Metode penelitian penggalian data atau *data mining* yang dipakai ialah metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), sehingga hasil penelitian ini bisa digunakan untuk menentukan kebijakan di daerah tersebut dan bisa mengurangi angka penyebaran Covid-19 dengan data terbaru yaitu tahun 2020 sampai 2022, dimana data diklasifikasikan dari *mean* terdekat, sehingga setiap kawasan di DKI Jakarta bisa diklasifikasikan menjadi kluster berdasarkan dampak

Covid-19 pada daerah tersebut, dan alokasi bantuan Covid-19 bisa ditetapkan berdasarkan informasi kluster tersebut, sehingga penindakan terhadap pandemi Covid-19 ini dapat dilaksanakan secara seefektif mungkin. Hasil dari klustering tersebut berupa visualisasi yang terdiri dari 3 kluster merah, kuning dan hijau, yang akan informatif. Selain itu untuk mencari dan menemukan indikasi penyebab peningkatan jumlah pasien positif covid maka pada penelitian dilanjutkan dengan melakukan pembuktiannya melalui *pearson correlation* untuk mengetahui apakah jumlah pasien yang belum melakukan vaksinasi dapat menyebabkan peningkatan jumlah pasien pada kluster merah. Dalam penelitian ini Korelasi Pearson untuk menentukan apakah ada korelasi antara data masyarakat yang belum melakukan vaksinasi dengan masyarakat yang terindikasi kasus positif covid, dengan itu dapat menghitung koefisien korelasinya. Juga dapat melihat hubungan antara populasi di suatu wilayah daerah DKI Jakarta akan berpengaruh untuk penyebab tingkat kasus positif Covid-19.

1.2. Batasan Masalah

Batasan masalah digunakan untuk membatasi penelitian yang dilakukan sehingga hanya fokus pada tujuan yang hendak dicapai. Batasan masalah yang digunakan yaitu:

1. Data yang digunakan dalam penelitian didapat dari website Corona Jakarta (<https://corona.jakarta.go.id/>)
2. Penggunaan data untuk penelitian ini adalah data kasus Covid-19 di Jakarta tahun 2020 – 2022 yang terdiri dari data jumlah yang terdeteksi positif dan kasus aktif. Perincian dataset yaitu Maret-Desember 2020, Januari-Desember 2021 dan Januari-November 2022.
3. Pendekatan kuantitatif digunakan sebagai teknik pengumpulan data dan *Purposive Sampling* digunakan untuk pengambilan sampel.
4. Kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) digunakan dalam metode penelitian *data mining*.
5. Variabel data dependen dan independen. Variabel independen merupakan variabel data-data yang digunakan untuk membuat klusterisasi yang direncanakan. Variabel ini akan memberikan pengaruh terhadap perubahan

pada variabel dependen. Variabel independent dalam penelitian ini yang terdiri dari: Positif, Dirawat, Sembuh, Meninggal dan Isoman.

6. Variabel dependen merupakan luaran klasterisasi yang dihasilkan dari model data mining yang dibuat, variabel ini dipengaruhi oleh variabel independent. Variabel dependen yang digunakan dalam penelitian ini yang terdiri dari: klaster merah untuk mengindikasikan daerah yang rawan dengan jumlah pasien Covid-19 terbanyak. Klaster kuning untuk mengindikasikan daerah yang cukup rawan dengan jumlah pasien Covid-19 hampir banyak. Klaster hijau untuk mengindikasikan daerah yang aman dengan jumlah pasien Covid-19 tidak terlalu banyak.
7. Algoritma *K-Means* dan *Fuzzy C-Means* digunakan untuk pembentukan model klasterisasi yang dihasilkan.
8. Penerapan data mining untuk pembentukan klasterisasi yang dihasilkan menggunakan alat bantu bahasa pemrograman Python dengan *platform Jupyter Notebook*.

1.3. Rumusan Masalah

Dari latar belakang yang dijelaskan, maka terdapat rumusan masalah yaitu:

Rumusan masalah pada penelitian ini diantaranya:

1. Bagaimana penerapan teknik data mining untuk pembentukan klasterisasi pasien pada kasus Covid-19 di DKI Jakarta?
2. Bagaimana melakukan implementasi algoritma *K-Means* dan *Fuzzy C-Means* dalam melakukan klastering pada kasus Covid-19 di DKI Jakarta?
3. Bagaimana hasil komparasi metode *K-Means* dan *Fuzzy C-Means* dalam melakukan *Klastering* pada kasus Covid-19 di Jakarta?

1.4. Tujuan Penelitian dan Manfaat Penelitian

1.4.1. Tujuan Penelitian

Tujuan dari penelitian ini diantaranya:

1. Untuk melakukan klastering pada kasus Covid-19 di DKI Jakarta.
2. Untuk melakukan implementasi algoritma *K-Means* dan *Fuzzy C-Means*

untuk klustering pada kasus Covid-19.

3. Untuk mengetahui dan membandingkan hasil penerapan algoritma *K-Means* dan *Fuzzy C-Means* dalam melakukan klusterisasi pada kasus Covid-19 di DKI Jakarta yang dihasilkan untuk membentuk klusterisasi pada kasus Covid-19 di DKI Jakarta.

1.4.2. Manfaat Penelitian

Manfaat dari penelitian yang dilakukan diantaranya:

1. Membantu pemerintah daerah provinsi DKI Jakarta dalam menentukan prioritas penanganan dan kebijakan masing-masing wilayah kota administratif di provinsi DKI Jakarta agar dapat menekan angka kenaikan kasus Covid-19.
2. Untuk melengkapi informasi yang tersedia di laman corona.jakarta.go.id dan memberikan kemudahan melakukan evaluasi penyebaran Covid-19 melalui visualisasi klusterisasi penyebaran Covid-19 di DKI Jakarta yang lebih informatif dengan menghasilkan 3 kluster merah, kuning dan hijau. mencerminkan jumlah kasus Covid-19 yang terjadi berdasarkan kluster jumlah pasien rawan dalam perawatan, sembuh, positif, isolasi mandiri, dan pasien meninggal,
3. Untuk dapat membantu pemerintah daerah DKI Jakarta dalam menentukan prioritas dan kebijakan masing-masing wilayah untuk menekan angka pertumbuhan kasus Covid-19 di DKI Jakarta.
4. Memberikan wawasan baru dari hasil penerapan algoritma *K-Means* dan *Fuzzy C-Means* untuk melakukan teknik klusterisasi pada kasus penyebaran Covid-19 di DKI Jakarta.

1.5. Sistematika Penulisan

Sistematika penulisan merupakan kerangka dari laporan skripsi yang dilakukan. Kerangka penulisan skripsi yaitu:

BAB I PENDAHULUAN

BAB I mendeskripsikan latar belakang masalah yang terdapat pada penelitian, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian.

BAB II TINJAUAN REFERENSI

BAB II mendeskripsikan dasar teori yang dipakai sebagai acuan dalam pembuatan penelitian ini, diperoleh dari berbagai buku, jurnal ilmiah, artikel serta referensi lainnya.

BAB III METODE PENELITIAN

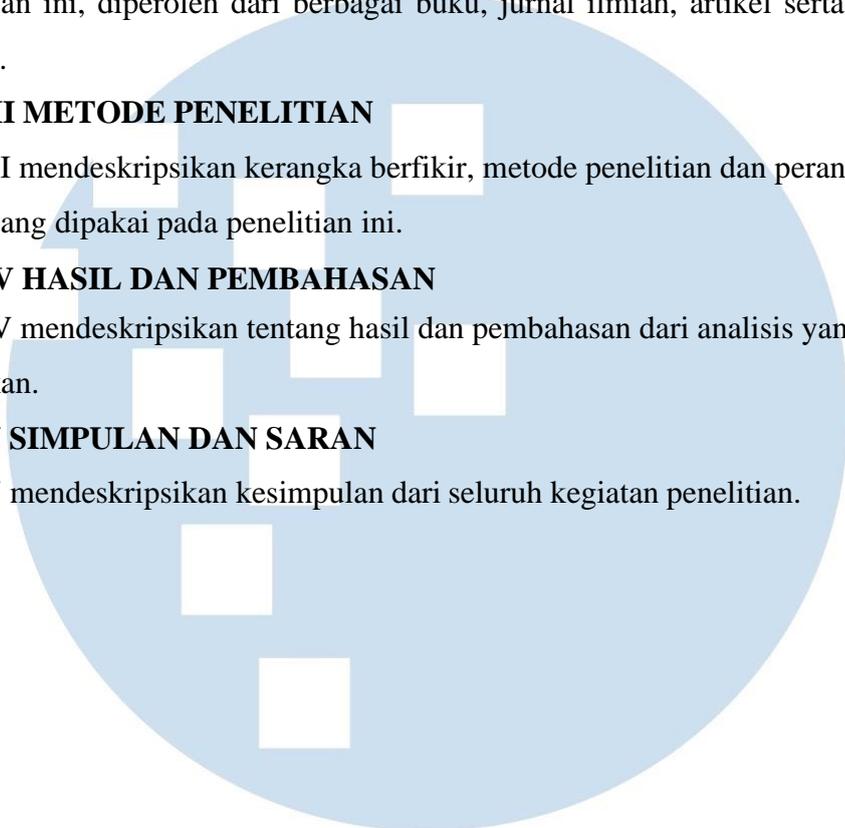
BAB III mendeskripsikan kerangka berfikir, metode penelitian dan perancangan UML yang dipakai pada penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

BAB IV mendeskripsikan tentang hasil dan pembahasan dari analisis yang sudah dilakukan.

BAB V SIMPULAN DAN SARAN

BAB V mendeskripsikan kesimpulan dari seluruh kegiatan penelitian.



UMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA