

BAB 2

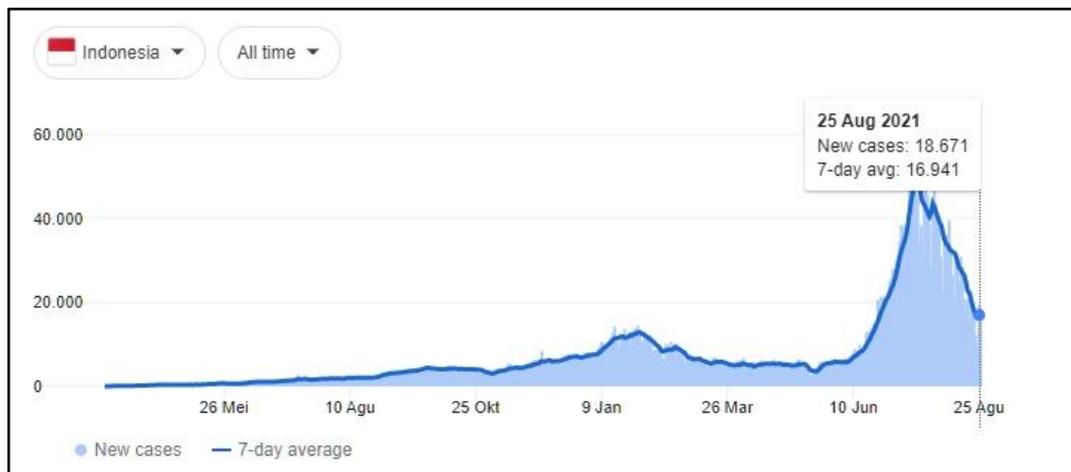
LANDASAN TEORI

2.1 COVID-19

COVID-19 adalah penyakit yang disebabkan oleh virus Severe Acute Respiratory Syndrome 2 (SARS-CoV-2). COVID-19 dapat menyebabkan gangguan sistem pernapasan, mulai dari gejala yang ringan seperti flu, hingga infeksi paru-paru, seperti pneumonia (World Health Organization, 2020).

Kasus pertama penyakit ini terjadi di Kota Wuhan, China, pada akhir Desember 2019. Setelah itu COVID-19 menular antar manusia dengan sangat cepat dan menyebar ke puluhan negara, termasuk Indonesia, hanya dalam beberapa bulan.

Menurut Gugus Tugas Penanganan COVID-19 (2020), di Indonesia, terhitung per tanggal 26 Agustus 2021, sebanyak 4.043.736 kasus terkonfirmasi, 3.669.966 orang sembuh dan 130.182 orang meninggal dunia. Dengan tingkat orang positif terkena COVID-19 sebesar 21,96% menjadikan Indonesia sebagai negara dengan kasus aktif tertinggi di Asia.



Gambar 2.1 Grafik Penambahan Kasus Harian COVID-19 di Indonesia

(Sumber: JHU CSSE COVID-19 Data, 2021)

Dikarenakan pandemi global yang terjadi banyak pihak berupaya ikut berperan serta untuk mengatasi, karena seperti yang diketahui, dunia seakan lumpuh mendadak sebab dari COVID-19. Para tenaga kesehatan pun angkat bicara

guna memberi himbauan kepada masyarakat untuk selalu mencuci tangan, menjaga jarak dan menggunakan masker ketika keluar rumah.

2.2 Data Mining

Data Mining merupakan metode pengolahan data berskala besar oleh karena itu *data mining* ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi (Turban, 2020). Secara umum, kajian data mining membahas metode-metode seperti *clustering*, klasifikasi, regresi, *variable selection*, dan *market basket* analisis.

Data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah sebagian besar data yang disimpan 10 dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. *Data mining* merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari basis data yang besar (Bastian, 2018).

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. Berawal dari beberapa disiplin ilmu, data mining bertujuan untuk memperbaiki teknik tradisional sehingga bisa menangani:

1. Jumlah data yang sangat besar
2. Dimensi data yang tinggi
3. Data yang heterogen dan berbeda sifat

Menurut para ahli, data mining merupakan sebuah analisa dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya (Jefri, 2013).

Adapun proses pengelompokan *data mining* berdasarkan tugas yang dapat dilakukan, yaitu:

1) *Description*

Secara sederhana, digambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan seringkali memberikan kemungkinan penjelasan untuk suatu pola.

2) *Estimation*

Estimasi mengarah ke variabel numerik. Nilai dari variabel target sebagai nilai prediksi disediakan menggunakan model yang dibangun. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

3) *Prediction*

Prediksi hampir sama dengan klasifikasi dan estimasi. Pada prediksi, terdapat proses untuk menemukan pola dari data dengan menggunakan beberapa variabel untuk melakukan prediksi variabel lain yang tidak diketahui nilainya.

4) *Classification*

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi untuk membedakan data ke dalam beberapa kelas. Klasifikasi melibatkan proses pemeriksaan karakteristik dari objek dan memasukkan objek ke salah satu kelas yang sudah didefinisikan.

5) *Clustering*

Pengklusteran merupakan pengelompokan *class* beberapa objek yang memiliki kemiripan. Pengklusteran bertujuan untuk mengelompokkan objek yang mirip satu sama lain dalam kelompok. Semakin besar kemiripan objek dalam suatu cluster dan tiap cluster maka semakin baik.

6) *Association*

Asosiasi dalam data mining adalah untuk menemukan atribut yang muncul dalam suatu waktu. Tugas asosiasi berusaha untuk mengungkap aturan untuk mengukur hubungan antara dua atau lebih atribut.

2.3 Clustering

Pada dasarnya *clustering* merupakan suatu metode dalam *data mining* untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik antara

satu data dengan data lainnya. *Clustering* terdapat 2 jenis yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa B, 2007).

Hierarchical clustering adalah metode pengelompokan data yang dimulai dengan melakukan pengelompokan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses dilanjutkan ke objek lain yang memiliki kedekatan kedua. Sedangkan pada *non-hierarchical clustering* pada dasarnya ditentukan terlebih dahulu nilai k yang diinginkan. Setelah nilai k ditentukan, baru proses *clustering* dilakukan, metode inilah yang disebut K-Means Clustering.

2.4 Algoritma K-Means Clustering

Salah satu metode *data mining* yang sering digunakan yaitu algoritma K-Means Clustering, untuk mengidentifikasi kelompok alami dari sebuah kasus yang didasarkan pada pengelompokan data yang memiliki kemiripan sehingga dengan pengelompokan tersebut dapat dilakukan analisis (Vulandari, 2017). Data-data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah *cluster*, di mana setiap *cluster* ini memiliki centroid atau titik pusat (Suntoro, 2019).

Berikut adalah tahapan-tahapan untuk menggunakan Algoritma K-Means:

- 1) Menentukan jumlah *cluster* (k) yang diinginkan pada *dataset*
- 2) Menentukan centroid secara *random* pada tahap awal

Rumus:

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}, \quad (2.1)$$

di mana:

V_{ij} adalah centroid/rata-rata *cluster* ke-i untuk variabel ke-j

N_i adalah jumlah data yang menjadi anggota *cluster* ke-i

i, k adalah indeks dari *cluster*

j adalah indeks dari variabel

X_{kj} adalah nilai data ke-k yang ada di dalam *cluster* tersebut untuk variabel ke-j

- 3) Menentukan jarak terdekat setiap data dengan centroid. *Euclidean distance* digunakan untuk menghitung jarak terdekat dengan centroid.

Rumus:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}, \quad (2.2)$$

Keterangan:

D_e adalah *Euclidean Distance*

i adalah banyaknya objek

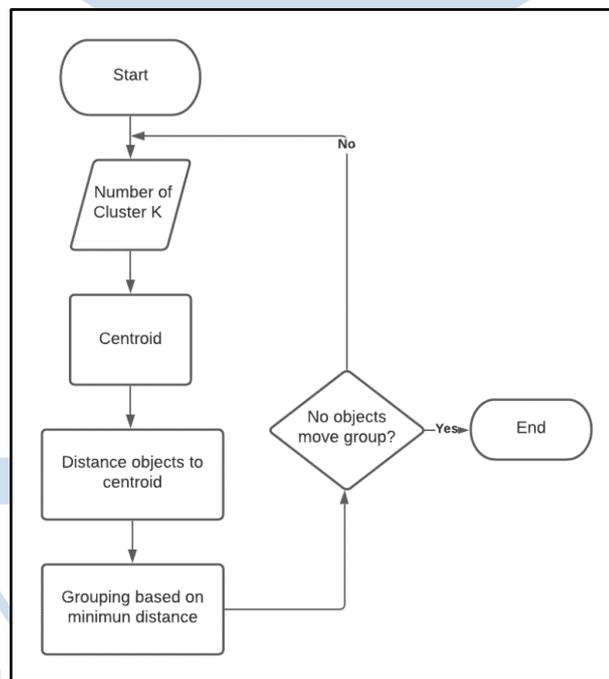
(x,y) merupakan koordinat objek

(s, t) merupakan koordinat centroid

- 4) Menghitung kembali *cluster* dengan keanggotaan *cluster* baru. Rata-rata dari semua data dijadikan sebagai pusat *cluster*. Proses berhenti ketika nilai pada *cluster* tidak berubah lagi.

Adapun cara kerja Algoritma K-Means Clustering dilakukan sampai proses iterasi stabil atau nilai centroid tetap yaitu:

- 1) Menentukan data centroid
- 2) Menghitung jarak antara centroid dengan masing-masing data
- 3) Mengelompokkan data berdasarkan jarak minimum



Gambar 2.2 *Flowchart* Algoritma K-Means Clustering

(Sumber: Hermawati, 2013)

2.5 Metode Elbow

Metode *Elbow* merupakan salah satu metode untuk menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik (Madhulatha, 2012). Metode ini memberikan gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Hasil persentase yang berbeda dari setiap nilai *cluster* ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Dikatakan nilai *cluster* tersebut paling baik jika pada grafik membentuk siku atau mengalami penurunan yang paling besar.

Untuk mendapatkan perbandingannya adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *cluster*. Jika semakin besar jumlah *cluster* k maka nilai SSE akan semakin kecil (Irwanto, Purwananto, & Soelaiman, 2012).

Berikut adalah tahapan-tahapan untuk melakukan Algoritma *Elbow* dalam menentukan nilai k pada Algoritma K-Means (Merliana, Ernawati, & Santoso, 2015):

- 1) Menginisialisasi awal nilai k
- 2) Menaikkan nilai k
- 3) Menghitung hasil *sum of square error* dari tiap nilai k
- 4) Melakukan analisis *sum of square error* dari nilai k yang mengalami penurunan secara drastis
- 5) Menetapkan nilai k yang berbentuk siku

Rumus:

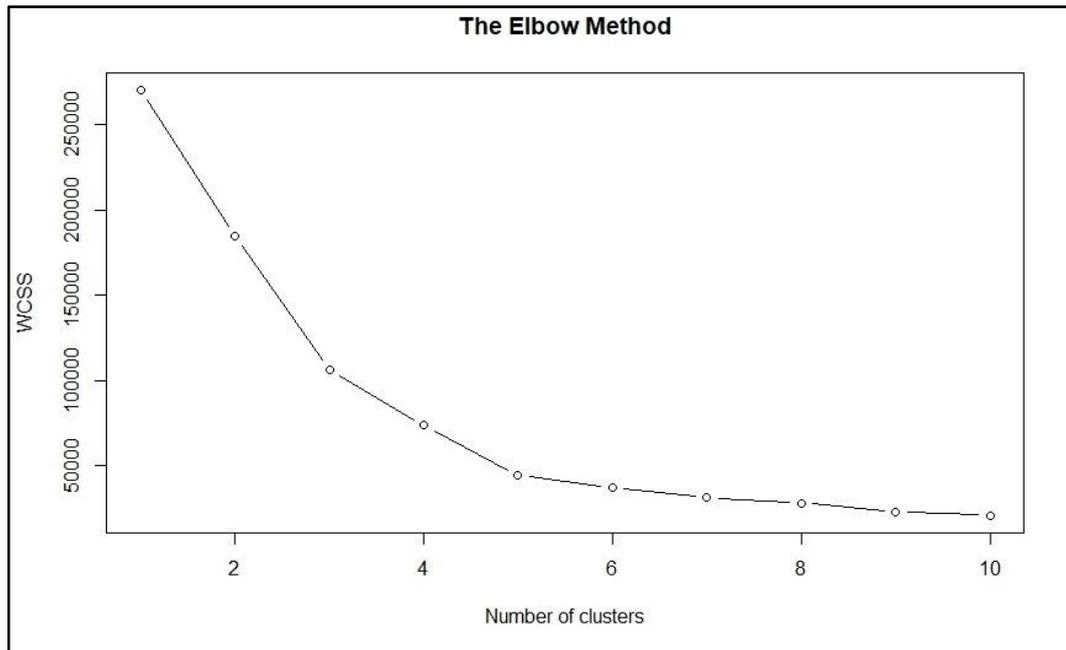
$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2, \quad (2.3)$$

Keterangan:

K adalah jumlah *cluster*

x_i adalah data ke- i

C_k adalah centroid *cluster*



Gambar 2.3 Gambar Metode *Elbow*

(Sumber: Herlambang, 2019)

2.6 Bahasa Pemrograman PHP

PHP adalah singkatan dari *Hypertext Preprocessor*. PHP adalah bahasa *scripting* yang dirancang khusus untuk mengembangkan situs berbasis *website*. Penggunaan umumnya adalah dengan pengambilan atau penarikan data dari *database SQL*. Bahasa pemrograman PHP menjalankan tugas pengembangan situs *website* umum.

Pada awalnya PHP memiliki singkatan dari frase *Personal Home Page*. PHP pertama kali diterapkan oleh Rasmus Lerdorf pada tahun 1994. Awalnya bahasa pemrograman PHP digunakan untuk halaman situs pribadi Lerdorf yang mana diperuntukkan untuk melacak pengunjung. Beberapa tahun kemudian, PHP dikenal setelah ditambahkan fungsi di dalamnya (Madcoms, 2016).

2.7 Database MySQL

Menurut Sibero (2013), “MySQL atau dibaca “My Sekuel” dengan adalah suatu RDBMS (*Relational Database Management System*) yaitu aplikasi sistem yang menjalankan fungsi pengolahan data”.

MySQL adalah sebuah manajemen basis data menggunakan perintah dasar SQL (*Structured Query Language*) yang cukup terkenal. MySQL adalah DBMS

(*Database Management System*) yang bersifat *open source* dengan dua bentuk lisensi, yaitu *free software* dan *shareware*.

SQL sendiri merupakan suatu bahasa yang dipakai di dalam pengambilan data pada *relational database*. Jadi MySQL adalah *database management system* yang menggunakan bahasa SQL sebagai bahasa penghubung antara perangkat lunak aplikasi dengan server *database*.

