

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada bulan Desember 2019 dunia digegerkan dengan penemuan virus yang mengerikan yaitu *Coronavirus* berjenis SARS-CoV-2 yang menimbulkan penyakit *Coronavirus Disease 2019* (COVID-19) di kota Wuhan, Tiongkok. Virus ini sangat berbahaya, karena penyebarannya sangat laju ke penjuru negara sehingga *World Health Organization* (WHO) menetapkan penyakit tersebut sebagai pandemi pada 12 Maret 2020. Di Indonesia, kasus positif pertama virus ini dikonfirmasi pada 3 Maret 2020.

Seiring meningkatnya penularan COVID-19 di Indonesia, banyak sekolah, perguruan tinggi, dan instansi baik pemerintah maupun swasta diliburkan sementara waktu. Beberapa pemerintah daerah bahkan menerapkan *lockdown* dan pemerintah pusat mengimbau masyarakat untuk melakukan karantina mandiri.

Dampak dari pandemi COVID-19 tidak hanya berimbas ke bidang kesehatan saja, tetapi juga berimbas hingga berbagai bidang kegiatan masyarakat. Masyarakat secara tidak langsung harus menggunakan media sosial untuk berkomunikasi jarak jauh saat melakukan karantina mandiri. Melalui media sosial daring, masyarakat dengan mudah membagikan dan mendapatkan informasi dari *social media* ke *social media* lainnya sehingga menjadi tren pembahasan. Maraknya penggunaan media sosial saat ini membuat masyarakat lebih sering mengutarakan opininya ke media sosial. Twitter merupakan media sosial yang banyak digunakan untuk mencurahkan opini tersebut.

Data opini yang dihasilkan Twitter itu penting dan bermanfaat bagi masyarakat atau organisasi jika diolah dan dianalisis. Banyaknya data opini yang didapatkan membuat data menjadi tidak berstruktur dan tidak berkelompok. Hal ini menimbulkan permasalahan berupa sulitnya mengetahui opini mana saja yang dapat masuk ke kelompok tertentu.

Maka dari itu, diperlukan program yang berfungsi untuk mengelompokkan data opini secara otomatis untuk mengetahui tren di masyarakat saat pandemi COVID-19. Penelitian ini bertujuan untuk mengelompokkan data opini ke suatu *cluster* atau kelompok dari Twitter. Dalam pengelompokan data, penulis menggunakan algoritma *K-Means Clustering*.

## **1.2 Perumusan Masalah**

Berlandaskan dari subbab latar belakang diatas, maka permasalahan yang akan distudi dalam penelitian ini adalah:

1. Bagaimana representasi dari data opini pengguna Twitter?
2. Bagaimana penerapan metode *clustering K-Means* dan *Davies Bouldin Indeks* (DBI) dalam mengelompokkan data opini.
3. Apa saja kata yang sering muncul berdasarkan masing-masing *cluster* dari hasil *clustering* pada data opini?
4. Apa saja visualisasi yang dapat ditampilkan dari hasil *clustering*?

## **1.3 Batasan Masalah**

Batasan masalah pada penelitian ini yakni:

1. Penelitian ini menggunakan data opini unggahan *tweet* berbahasa Indonesia dengan kata kunci “*virus corona*” atau “*covid-19*” dari pengguna Twitter yang diunggah mulai dari bulan Maret 2020 hingga April 2020.
2. Pengelompokan data opini menggunakan metode *clustering K-Means* dan *Davies Bouldin Index* sebagai metode penentu jumlah *cluster*.
3. Dalam mengelompokkan data opini digunakan bahasa pemrograman Python.
4. Visualisasi data hasil *clustering* menampilkan kata yang sering muncul pada tiap *cluster* dalam bentuk *word cloud*.

#### **1.4 Tujuan Penelitian**

Berlandaskan pada subbab perumusan masalah diatas, tujuan yang hendak diperoleh dari penyusunan penelitian ini antara lain:

1. Mengelompokkan data opini pengguna media sosial Twitter berdasarkan parameter yang ada.
2. Mewujudkan program yang dapat menyuguhkan hasil data setelah melalui tahap *clustering* beserta visualisasinya.

#### **1.5 Manfaat Penelitian**

Apabila program dapat diterapkan, maka manfaat penelitian ini adalah:

1. Memberikan informasi mengenai pembahasan masyarakat saat pandemi COVID-19 dan bagaimana pengelompokan data opini.
2. Sebagai bahan pertimbangan bagi Pemerintah Pusat dan Dinas atau Instansi terkait dengan infrastruktur dalam pengambilan keputusan maupun

merumuskan kebijakan/program yang tepat dalam rangka menanggulangi bencana pandemi COVID-19.

## **1.6 Sistematika Penulisan**

Tujuan sistematika penulisan adalah untuk menyederhanakan pada pemahaman alur pemikiran keseluruhan tugas akhir. Secara garis besar tugas akhir ini terdiri dari lima bab dengan sistematika sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan, manfaat, dan sistematika penulisan. Pada pendahuluan ini akan diuraikan mengenai hal-hal apa yang menjadi alasan penulisan laporan penelitian tugas akhir ini.

### **BAB II TINJAUAN PUSTAKA**

Dalam bab ini memaparkan penjelasan tinjauan mengenai hal-hal yang berhubungan dengan penerapan algoritma *text mining* yaitu *k-means clustering* untuk mengelompokkan data opini dan yang mendukung laporan penelitian. Teori-teori tersebut akan ditinjau dan digunakan sebagai pedoman untuk mendukung penulisan laporan penelitian.

### **BAB III METODE PENELITIAN**

Bab ini membahas berbagai macam pendekatan yang digunakan dalam penelitian yang terdiri dari pengumpulan data, *text preprocessing*, TF-IDF,  *Davies Bouldin index*, dan implementasi algoritma *K-Means*

#### **BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Bab ini menjabarkan hasil penelitian dari metode yang digunakan. Bab ini juga menyajikan visualisasi informasi sebagai sarana interaksi pengguna dengan tampilan hasil mengekstraksi pola informasi.

#### **BAB V KESIMPULAN DAN SARAN**

Bab ini mengutarakan kesimpulan yang didapatkan dari hasil dan pembahasan pada bab keempat serta saran mengenai permasalahan yang dapat dikembangkan selanjutnya.