

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Sejak awal tahun 2020, ramai di masyarakat lingkup dunia tentang adanya sebuah virus mematikan yang dapat menyebar luas dalam waktu yang singkat bernama SARS-CoV-2. SARS-CoV-2 ini merupakan jenis virus terbaru dari coronavirus. Penyakit yang disebabkan oleh virus ini dinamakan *Coronavirus disease 2019 (COVID-19)*. Berdasarkan data yang diambil dari *World Health Organization (WHO)*, terdapat sebanyak 4,347,935 kasus yang terjadi di seluruh dunia, dengan 79,187 kasus baru per tanggal 15 Mei 2020 [1]. Dengan level percepatan penyebaran yang begitu tinggi dan belum ditemukannya vaksin untuk COVID-19, menyebabkan terjadinya kekacauan ditengah masyarakat. Hal ini mempengaruhi banyak sektor kehidupan, dari ekonomi, politik, industri [2], pendidikan, medis, dan juga kehidupan sosial.



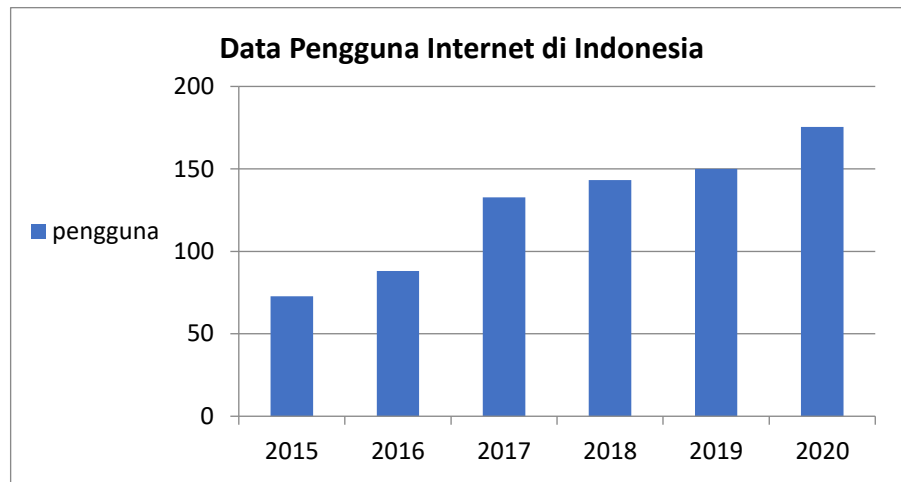
Gambar 1.1 Jumlah Kasus Terkonfirmasi COVID-19 [3]

Dampak dari terjadinya pandemi ini membuat banyak kekacauan di berbagai belahan dunia, tidak terkecuali Indonesia. Awal bulan Maret 2020, Presiden Joko Widodo menyampaikan langsung temuan kasus COVID-19 pertama di Indonesia [4]. Total kasus terkonfirmasi pada bulan Maret yakni 1.528 kasus. Kemudian pada bulan selanjutnya jumlah kasus terkonfirmasi mengalami

peningkatan sebanyak 8.590, yang berarti pada bulan April terdapat 10.118 kasus terkonfirmasi. Selanjutnya pada bulan Mei terdapat sebanyak 26.473 kasus terkonfirmasi. Pada bulan Juni terdapat sebanyak 56.385 kasus terkonfirmasi COVID-19, dan untuk data terbaru yakni bulan Juli jumlah kasus terkonfirmasi meningkat sangat tajam dari jumlah sebelumnya. Pada bulan Juli terjadi peningkatan sebanyak 51.991, total kasus terkonfirmasi yaitu 108.376 [3].

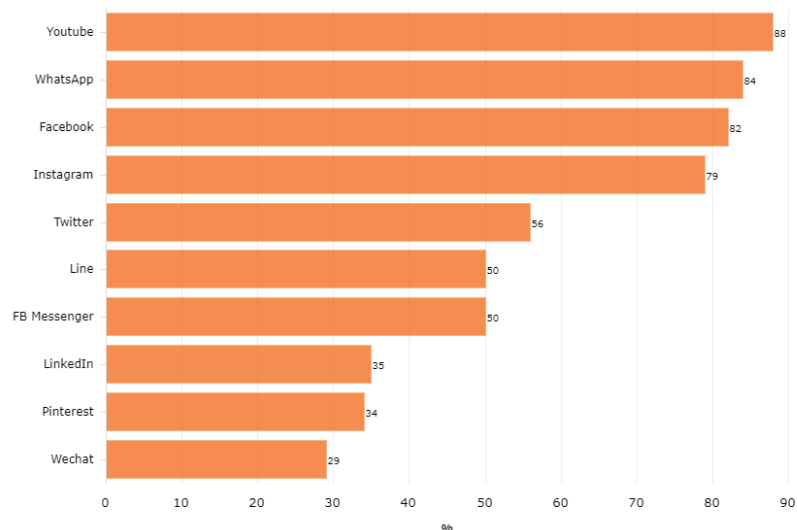
Banyak penutupan sekolah, fasilitas umum, pembatasan moda transportasi, pelayanan masyarakat, dan juga berbagai macam hal lainnya. Di Indonesia sendiri, pemerintah mengeluarkan kebijakan Pembatasan Sosial Berskala Besar (PSBB). PSBB meliputi pembatasan kegiatan pendidikan, pembatasan kegiatan kerja (dengan diberlakukannya *Work From Home* oleh berbagai instansi), pembatasan kegiatan beribadah, larangan masyarakat untuk berkumpul, pembatasan atau larangan bepergian keluar kota, dan lain sebagainya. Kebijakan PSBB ini diatur dalam Peraturan Pemerintah No. 21 Tahun 2020 tentang Pembatasan Sosial Berskala Besar (PP PSBB) dalam Rangka Percepatan Penanganan Corona Virus Disease (COVID-19) [5].

Indonesia merupakan salah satu negara dengan populasi pengguna internet terbesar di dunia. Berdasarkan sumber dari *We are Social* pada tahun 2020, Indonesia mencapai angka 175,4 juta orang mengakses internet. Hal ini menunjukkan adanya pertumbuhan populasi mengakses internet sebesar 17 persen dalam satu tahun terakhir, yang berarti ada 25,3 juta mengakses internet baru sejak tahun 2019 [6].



Gambar 1.2 Data Pengguna Internet Dalam Satuan Juta [7]

*Twitter* merupakan salah satu platform media sosial yang paling banyak digunakan oleh masyarakat Indonesia. Menurut sumber *We are Social* dan *Hootsuite* tahun 2020, *Twitter* menempati posisi kelima pada kategori media sosial yang sering digunakan dengan jumlah presentase pengguna sebesar 56% setelah *Youtube*, *Whatsapp*, *Facebook* dan juga *Instagram*.



Gambar 1.3 Media Sosial yang Sering Digunakan di Indonesia [7]

*Twitter* banyak digunakan oleh masyarakat Indonesia dikarenakan penggunaannya yang tergolong cukup mudah. Untuk melakukan pendaftaran, pengguna hanya membutuhkan *e-mail* dan juga nomor ponsel. Salah satu fitur dari *Twitter* yang paling sering digunakan yaitu fitur *tweet*. Dengan menggunakan fitur ini pengguna dapat menuliskan pemikiran, pendapat serta opininya. *Tweet-tweet* yang ditulis oleh para pengguna *platform* ini tentu dapat diolah menjadi satu informasi yang berguna.

Setiap individu memiliki pendapat dan opini yang berbeda-beda. Pendapat ini sangat penting dan juga merupakan salah satu hal yang mempengaruhi perilaku utama manusia. Sentimen analisis (*sentiment analysis*) atau yang juga sering disebut sebagai *opinion mining* merupakan bidang studi yang menganalisa opini masyarakat, sentimen, evaluasi, penilaian, sikap, dan emosi terhadap sebuah produk, pelayanan, organisasi dan perhimpunan, seorang tokoh, dan isu atau masalah serta peristiwa yang terjadi pada masyarakat itu sendiri [8]. Sentimen analisis banyak digunakan di berbagai bidang dalam kehidupan sehari-hari. Seperti misalnya dalam bidang bisnis, sentimen analisis digunakan untuk memprediksi harga saham, menganalisis kebutuhan pasar, hingga penyusunan strategi pemasaran berdasarkan sentimen masyarakat. Sentimen analisis juga digunakan untuk mengetahui opini publik terkait isu, kejadian, dan juga peristiwa yang terjadi di masyarakat.

Dalam penelitian ini akan dibahas mengenai sentimen analisis masyarakat terhadap pandemi COVID-19 pada media sosial *Twitter*. Berdasarkan peristiwa yang saat ini sedang ramai di masyarakat, banyak pengguna media sosial yang memberikan opini, pendapat, serta pemikirannya terhadap COVID-19 pada *platform* media sosial *Twitter*. Hal ini menarik untuk diteliti guna mengetahui opini masyarakat tentang pandemi yang sedang terjadi sekarang ini. Untuk menunjang penelitian tentang hal tersebut, dibutuhkan algoritma untuk mengklasifikasikan komentar masyarakat di media sosial *Twitter*, baik itu komentar positif maupun komentar negatif.

Ada beberapa algoritma yang dapat digunakan untuk klasifikasi komentar pada media sosial *Twitter*, diantaranya adalah *Naïve Bayes* [9][10], *K-Nearest*

*Neighbor* [11][12], *Decision Tree* [13][14], dan juga *Support Vectore Machine* [13][15].

Algoritma *Naïve Bayes* merupakan salah satu algoritma klasifikasi dengan metode probabilitas dan statistik yang pertama kali diciptakan oleh ilmuwan dari Inggris bernama *Thomas Bayes*. Algoritma ini memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal dengan nama Teorema Bayes [16]. Kelebihan dari algoritma ini yaitu sangat sederhana, cepat, mudah, dan kinerjanya yang cukup baik. Algoritma ini juga memiliki beberapa kekurangan seperti dibutuhkan banyak perhitungan, banyaknya celah untuk mengurangi keefektifan metode ini dan akibatnya meloloskan dokumen ke dalam kelas tertentu padahal jelas-jelas dokumen tersebut tidak layak berada di kelas tersebut [17].

Algoritma *K-Nearest Neighbor* merupakan salah satu algoritma yang paling sering digunakan untuk klasifikasi. Algoritma ini merupakan salah satu metode yang menerapkan algoritma supervised [18]. Akurasi algoritma ini ditentukan oleh ada dan tidaknya data yang tidak relevan, atau jika bobot fitur tersebut setara dengan relevansinya terhadap klasifikasi. Algoritma ini memiliki beberapa kelebihan yaitu efektif saat digunakan untuk data dengan jumlah yang besar dan mampu menghasilkan data yang jelas serta cukup kuat [19]. Selain kelebihan, algoritma *K-Nearest Neighbor* juga memiliki kekurangan yaitu prediksi yang diberikan dapat mengandung banyak *noise* ketika data memiliki banyak *noise* [20].

Algoritma *Decision Tree* merupakan salah satu metode klasifikasi *data mining* yang paling populer digunakan. Algoritma ini digunakan untuk menangani kasus-kasus dengan karakteristik data dinyatakan dengan pasangan atribut serta nilainya, keluaran data yang biasanya dalam bentuk nilai diskrit, serta data yang memiliki *missing value* [19]. Kelebihan algoritma *Decision Tree* yakni dapat dengan mudah menangani interaksi fitur yang tidak menggunakan parameter dan mudah diinterpretasi. Algoritma *Decision Tree* ini juga memiliki kekurangan yaitu tidak mendukung adanya *online learning* dan juga harus membangun ulang pohon keputusan apabila terdapat data yang baru [21].

Algoritma *Support Vector Machine (SVM)* merupakan metode untuk melakukan prediksi baik klasifikasi maupun regresi. Algoritma ini memiliki prinsip dasar *linier classifier* atau klasifikasi yang secara linier dapat dipisahkan. Akan tetapi algoritma ini telah dikembangkan supaya dapat menangani problem *non-linier* dengan measukan konsep kernel pada *workspace* yang memiliki dimensi tinggi [19]. Kelebihan algoritma *Support Vector Machine* ini adalah memiliki tingkat akurasi yang tinggi, dan baik secara teoritis dengan kernel yang tepat. Adapula kekurangan dari algoritma ini yaitu sangat rumit untuk diimplementasikan dan juga membutuhkan biaya yang cukup besar [21].

Berdasarkan hasil perbandingan dari algoritma klasifikasi yang ada, algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dinilai cocok untuk digunakan dalam pengklasifikasian teks komentar karena kedua algoritma tersebut sederhana, tingkat akurasi yang didapatkan rata-rata tinggi, dan juga fleksibel. Jadi pada penelitian ini, akan dilakukan perbandingan antara dua algoritma yaitu *Naïve Bayes* dan *K-Nearest Neighbor* dalam klasifikasi sentimen masyarakat terhadap COVID-19 di media sosial Twitter.

Tujuan yang ingin dicapai dalam penelitian ini adalah terciptanya sebuah model *sentiment analysis* untuk topik COVID-19 pada media sosial *Twitter*.

## 1.2 Rumusan Masalah

Pandemi COVID-19 yang terjadi sekarang menimbulkan banyak kekacauan dalam berbagai bidang dan tentunya menciptakan keresahan yang bergulir di tengah masyarakat. Sehingga dibutuhkan model *sentiment analysis* untuk klasifikasi komentar guna mengetahui sentimen dan opini publik terhadap peristiwa yang sedang terjadi ini. Dari rumusan masalah tersebut, maka munculah pertanyaan penelitian sebagai berikut:

- a. Seberapa besar hasil akurasi klasifikasi komentar menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam opini COVID-19 pada media sosial *Twitter*?
- b. Seberapa akurat sistem dapat memberikan prediksi kepada data pengujian?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, tujuan penelitian ini adalah sebagai berikut:

- a. Implementasi algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen masyarakat terhadap COVID-19 pada media sosial *Twitter*.
- b. Mendapatkan hasil akurasi dari algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.

### 1.4 Manfaat Penelitian

Berdasarkan dari tujuan penelitian yang sudah disebutkan, adapun manfaat penelitian ini adalah sebagai berikut:

- a. Manfaat bagi penulis, dapat menerapkan ilmu dalam bidang ilmu komputer NLP.
- b. Manfaat bagi akademik, diharapkan dapat memberikan kontribusi dalam bidang *healthcare* dan keilmuan NLP.
- c. Manfaat bagi kalangan umum, diharapkan dapat memberikan informasi tentang perkembangan virus dan analisis data.

### 1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

- a. Penelitian ini menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.
- b. Penelitian ini menggunakan metode klasifikasi.
- c. Menggunakan data dari media sosial *Twitter* dengan *keyword* COVID19.