

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan teknologi dan penyebaran informasi di internet selalu meningkat dari waktu ke waktu. Teknologi informasi adalah salah satu hal yang tidak akan lepas dari kehidupan manusia. Manusia akan kesusahan dalam berkomunikasi dan menyampaikan informasi tanpa adanya teknologi. Berita merupakan media informasi yang juga turut mengalami peningkatan. Pada awalnya banyak lembaga penyaluran informasi menyampaikan berita melalui media cetak koran beralih ke media elektronik seperti radio dan televisi. Seiring berjalannya waktu, berita memiliki media penyebaran baru yang akan mendukung mobilitas manusia yang semakin tinggi, yaitu media digital menggunakan sistem berbasis *web* secara *update*. Berdasarkan laporan DataReportal, per Januari 2022, pengguna internet di Indonesia mencapai 204,7 juta pengguna. Jumlah ini mencakup 73,7% dari total populasi Indonesia. Jumlah pengguna ini meningkat tipis 1,03% dibandingkan dengan tahun sebelumnya pada Januari 2021 yaitu tercatat sebanyak 202,6 juta pengguna (DataReportal, 2022).

Salah satu media digital yang menggunakan sistem berbasis *web* adalah portal berita detik.com. Detik.com adalah sebuah portal *web* yang berisikan artikel dan berita daring di Indonesia yang dapat diakses pada alamat URL [www.detik.com](http://www.detik.com). Berita dan artikel yang disampaikan terdiri dari beberapa kategori seperti kesehatan, politik, olahraga, teknologi, dan masih banyak lagi kategori berita lainnya. Beberapa tahun belakangan ini, banyak berita yang dicari oleh masyarakat dengan kategori kesehatan terutama berita mengenai pandemi Covid-19 yang sempat menggemparkan Indonesia hingga dunia belakangan ini.

Berita mengenai Covid-19 (*Corona Virus Disease 19*) beberapa tahun belakangan ini sangat ramai diperbincangkan dari awal kehadirannya di Wuhan, China hingga menyebar ke seluruh dunia termasuk Indonesia dan menelan banyak korban

jiwa. Berdasarkan data dari laman covid19.go.id, dari awal ditemukan hingga pada 01 April 2022 Covid-19 telah memakan korban jiwa di Indonesia hingga mencapai 155.164 orang meninggal (Covid19.go.id, 2022). Berita yang mengangkat tema Covid-19 pun banyak tersebar lewat berbagai *platform* media, dari media massa hingga media sosial. Hal tersebut memudahkan masyarakat untuk mendapatkan informasi tentang virus tersebut. Namun, dengan banyaknya berita yang beredar, banyak oknum-oknum tidak bertanggung jawab yang membuat dan menyebarkan berita-berita *hoax* terutama di media sosial, sehingga membuat kebanyakan orang semakin merasa was-was dan khawatir hingga menimbulkan perasaan cemas dan panik akan virus ini, yang mana hal tersebut tentu dapat mempengaruhi imun tubuh dari orang tersebut. Oleh karena itu, dibutuhkan model *sentiment analysis* untuk mengklasifikasikan berita Covid-19 pada portal berita detik.com menjadi data sentimen positif, netral, dan negatif. Berita dianggap positif apabila terdapat kalimat-kalimat bersifat positif, kalimat dianggap positif apabila terdapat kata-kata yang mengandung makna positif. Berita dianggap netral apabila terdapat kalimat-kalimat bersifat netral, kalimat dianggap netral apabila terdapat kata-kata yang mengandung makna netral. Berita dianggap negatif apabila terdapat kalimat-kalimat bersifat negatif, kalimat dianggap negatif apabila terdapat kata-kata yang mengandung makna negatif.

*Sentiment analysis* atau *opinion mining* merupakan proses mengemukakan informasi dengan mengklasifikasi dokumen teks ke dalam beberapa kelompok yang sesuai dengan keseluruhan sentimen yang diterangkan di dalam setiap dokumen tersebut (P. Setiawan, 2018). Pengimplementasian sistem *sentiment analysis* dapat digunakan untuk menganalisis sentimen dari data teks salah satunya yaitu teks berita. Isi berita yang dipublikasi dapat memunculkan opini berita positif, negatif, maupun netral terhadap suatu hal yang sedang dibahas oleh masyarakat. Terutama berita tentang Covid-19 yang sangat berpengaruh pada kesehatan mental dan fisik seseorang. Beberapa algoritma yang dapat digunakan untuk mengklasifikasi berita Covid-19 pada portal berita detik.com, salah satunya adalah algoritma *K-Nearest Neighbor*.

Algoritma *K-Nearest Neighbor* (K-NN) adalah salah satu algoritma yang paling sering kali digunakan untuk klasifikasi. Algoritma K-NN adalah salah satu metode yang menerapkan algoritma *supervised*. Akurasi algoritma K-NN ditentukan oleh ada dan tidaknya data yang tidak relevan atau jika bobot fitur tersebut sebanding dengan relevansinya terhadap klasifikasi. Kelebihan dari algoritma K-NN ini yaitu efektif saat dipakai untuk data dengan jumlah yang besar dan sanggup menghasilkan data yang cukup kuat dan jelas (Fairuz, 2020). Oleh karena itu, algoritma K-NN ini sangat cocok digunakan untuk mengklasifikasi berita, karena berita lebih banyak memuat kata yang lebih banyak dibandingkan dengan satu postingan di *Facebook* dan sebuah *Twitter*.

Penelitian lain yang juga menggunakan metode yang sama juga pernah dilakukan oleh Ar Razi dengan judul “Klasifikasi Penerimaan Beasiswa Aceh Carong (Aceh Pintar) di Universitas Malikussaleh Menggunakan Algoritma KNN (*K-Nearest Neighbors*)”, dengan hasil algoritma K-Nearest Neighbor cukup efektif dan efisien dalam mengklasifikasikan penerima beasiswa Aceh Carong dengan hasil 82,00% (Razi, 2022).

Penelitian serupa juga pernah dilakukan oleh Faisal Briliansyah dengan judul penelitian “Sistem Klasifikasi Kategori Berita Menggunakan Metode *K-Nearest Neighbor*” dengan hasil presentase nilai *accuracy* sebesar 76%, *precision* 35%, *recall* 35%, *f-measure* 35%, *specificity* 84%, dan UAC 60% (Briliansyah, 2020).

Berdasarkan latar belakang di atas, maka penulis mengangkat judul “**Analisis sentimen berita Covid-19 pada portal berita detik.com menggunakan algoritma *K-Nearest Neighbor***”. Adapun penelitian ini bertujuan untuk mengklasifikasikan berita Covid-19 pada portal berita detik.com menjadi kelompok positif, negatif, dan netral menggunakan metode algoritma *K-Nearest Neighbor*. Hasil dari penelitian ini akan memberikan gambaran kepada masyarakat umum apakah berita Covid-19 pada portal berita detik.com cenderung ke berita positif, negatif, atau netral.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana mengimplementasikan teknik *sentiment analysis* terhadap berita Covid-19 pada portal berita detik.com menggunakan metode algoritma *K-Nearest Neighbor* ?
2. Bagaimana tingkat akurasi sistem dengan metode *K-Nearest Neighbor* dalam penelitian *sentiment analysis* terhadap berita Covid-19 ?

## 1.3 Batasan Masalah

Agar tujuan dari penelitian ini tercapai, maka penelitian ini perlu dibatasi. Adapun batasan penelitian yang dibuat penulis yakni sebagai berikut:

1. Metode yang digunakan pada penelitian *sentiment analysis* ini adalah metode algoritma *K-Nearest Neighbor*.
2. Data sumber penelitian yang digunakan didapat dari portal berita detik.com.
3. Data yang diambil hanya berita mengenai Covid-19.
4. Data yang diambil sebanyak 450 data untuk data latih dan 50 data untuk data uji dalam kurun waktu dari tahun 2020-2021 dengan menggunakan *crawling* berdasarkan *term* Covid-19.

## 1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian ini yaitu sebagai berikut:

1. Menerapkan teknik *sentiment analysis* untuk mendapatkan informasi berupa klasifikasi yang dihasilkan dari setiap berita Covid-19 menggunakan metode *K-Nearest Neighbor*.
2. Melihat sejauh mana algoritma *K-Nearest Neighbor* dalam mengenali pola pada sebuah berita untuk mengetahui klasifikasi berita Covid-19.
3. Menganalisis sejauh mana tingkat klasifikasi positif, negatif, dan netral yang dihasilkan dari sebuah berita Covid-19 pada portal berita detik.com.

4. Mengetahui tingkat akurasi sistem dengan metode *K-Nearest Neighbor* dalam penelitian *sentiment analysis* terhadap berita Covid-19.

### **1.5 Manfaat Penelitian**

Penelitian diharapkan dapat membawa manfaat. Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui sejauh mana keakuratan metode algoritma *K-Nearest Neighbor* untuk diterapkan pada penelitian *sentiment analysis* berita Covid-19 pada portal berita detik.com.
2. Sebagai studi pustaka pada kegiatan-kegiatan penelitian selanjutnya.
3. Diharapkan dapat menjadi suatu referensi yang berguna bagi dunia akademik untuk mengetahui sejauh mana tingkat kemampuan algoritma *K-Nearest Neighbor* dalam melakukan analisis teks pada sebuah berita.