

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Perkembangan internet telah mengubah cara manusia memperoleh dan mengumpulkan informasi. Digitalisasi informasi membuat manusia makin beralih pada media penyimpanan elektronik. Keuntungan utama dari media penyimpanan elektronik adalah kemudahan dalam menyimpan dan mengakses informasi (*information retrieval*). Kemudahan ini pada akhirnya membuka peluang baru, yaitu "Bagaimana cara mendapatkan pengetahuan dari kumpulan informasi yang ada?".

Proses mendapatkan pengetahuan dari data kumpulan data ini dikenal sebagai *knowledge discovery in database* (KDD) atau yang sering disebut *data mining*. *Data mining* dilakukan pada data yang terstruktur. Data tersebut umumnya disimpan dalam *database* dengan struktur dan tipe data yang dapat dikenali oleh komputer. Namun, internet umumnya menyimpan data dalam format teks yang tidak langsung dapat dikenali oleh komputer (Shu & Ye, 2023).

*Text mining* adalah metode khusus *data mining* yang digunakan untuk memperoleh pengetahuan dari data teks yang tidak terstruktur (S, 2021). Proses *text mining* melibatkan ekstraksi informasi melalui pengenalan dan penjelajahan pola dalam data teks. Salah satu metode yang sering digunakan dalam *text mining* adalah *Naïve Bayes* (Chintalapudi dkk., 2021).

Algoritma *Naïve Bayes* sering digunakan oleh peneliti untuk membangun sistem pengklasifikasian. Sebagai contoh, (Donnellan dkk., 2022) menerapkan metode *Naïve Bayes* untuk membedakan antara 'definisi keingintahuan' dan 'minat' pada seseorang. Hasil pengklasifikasi ini digunakan untuk memberikan dasar kejelasan konseptual yang sangat dibutuhkan untuk memfasilitasi penelitian mengenai psikologis manusia di masa depan. Klasifikasi menggunakan algoritma *Naïve Bayes* juga digunakan oleh (Alamer, 2023) untuk memprediksi penyakit

berdasarkan kumpulan data medis. Hasil dari penelitian ini digunakan untuk membangun sistem informasi kesehatan cerdas.

Penulis juga menerapkan metode *Naïve* untuk menganalisis sentimen publik. Analisis sentimen atau *opinion mining* merupakan *emotional artificial intelligence* yang mengacu pada pemanfaatan pemrosesan bahasa alami, *text mining*, dan linguistik komputasional, untuk mengenali, mengevaluasi, dan memeriksa keadaan emosional secara metodis (Alsaeedi & Khan, 2019).

Penelitian ini dilakukan pada media *online* Twitter untuk mengukur polaritas sentimen terhadap “Angka Prevalensi Stunting di Indonesia. Mengklasifikasikan twit ke dalam beberapa kelas “positif, negatif, dan netral” dan hasil klasifikasi tersebut berguna untuk memahami pandangan, sikap, dan perasaan masyarakat (Xu dkk., 2022) terhadap program pencegahan dan penurunan angka prevalensi stunting, serta memetakan opini publik sebagai dasar pembentukan sistem rekomendasi berbagai kebijakan kesehatan dan sosial bagi pemerintah.

Stunting merupakan kondisi gagal tumbuh pada anak balita akibat kekurangan gizi kronis terutama pada 1.000 Hari Pertama Kehidupan (HPK). Stunting yang merupakan kegagalan tumbuh yang mengacu pada anak yang terlalu pendek dibandingkan usianya, dijadikan salah satu indikator target ketercapaian gizi pada anak (UNICEF/WHO/WORLD BANK, 2021).

Anak stunting membawa risiko lebih tinggi menderita penyakit kronis di masa dewasanya. Peningkatan resiko penyakit degeneratif, penurunan fungsi reproduksi, penurunan kapasitas dan prestasi belajar, serta penurunan produktifitas kerja. Sehingga berkonsekuensi pada kualitas hidup seorang anak (WHO, 2018). Hal tentu akan memengaruhi bidang ekonomi dengan peningkatan biaya kesehatan yang dikeluarkan. Stunting dan malnutrisi diperkirakan berkontribusi pada berkurangnya 2-3% Produk Domestik Bruto (PDB) setiap tahunnya (Akseer dkk., 2022).

Analisis sentimen dengan pengklasifikasian menggunakan metode *Naïve Bayes Classifier* (NBC) menjadi penting untuk melihat sejauh mana seruan berbagai program yang berkaitan dengan stunting mampu menjangkau masyarakat Indonesia dan sejauh mana Twitter mampu menjadi diseminasi informasi kesehatan di Indonesia. Hal ini menjadi landasan mendasar bagi penulis untuk melakukan penelitian tugas akhir dengan judul “Analisis Sentimen Publik Terhadap Program Penurunan Angka Prevalensi Stunting Indonesia Menggunakan Data Twitter Dengan Metode *Naïve Bayes*”.

## 1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan beberapa masalah yang menjadi inti dari penelitian ini, yaitu:

1. Bagaimana menarik data teks dari kumpulan twit twiter dan menjadikannya data bersih?
2. Bagaimana menerapkan algoritma *Naïve Bayes Classifier* (NBC) dalam proses *text mining*, sehingga didapatkan klasifikasi sentimen dari data twitter?
3. Bagaimana memvualisasikan hasil pengklasifikasian data teks dalam memantau perilaku interaksi masyarakat terhadap program-program pemerintah di Indonesia?

## 1.3 Batasan Masalah

Dengan luasnya cakupan yang dapat terkait dengan penelitian ini dan untuk keseragaman pemahaman dalam penelitian, maka terdapat batasan-batasan yang perlu diberlakukan pada penelitian ini. Batasan-batasan tersebut adalah:

1. Penelitian ini hanya mengambil sampel data twit Twitter berbahasa Indonesia yang berkenaan dengan pembahasan tentang stunting.
2. Pengklasifikasian data teks pada penelitian ini menggunakan metode *Naïve Bayes Classifier* (NBC).
3. Penelitian ini mengklasifikasikan twit menjadi tiga kategori, positif, negatif, dan netral.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah dan batasan masalah yang telah diuraikan, maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis sentimen publik berbasis data Twitter terhadap angka prevalensi stunting di Indonesia dengan menggunakan metode klasifikasi *Naïve Bayes*.
2. Mendapatkan data bersih untuk dilakukan pemrosesan data (*text analytic/text mining*).

3. Memahami metoda *Naïve Bayes* dalam *text analytic/text mining* untuk pengklasifikasian data teks.
4. Menganalisis penerapan metode *Naïve Bayes* pada sentimen publik.
5. Merancang sistem klasifikasi teks.
6. Memvualisasikan hasil pengklasifikasian data teks dalam memantau perilaku interaksi masyarakat terhadap program-program pemerintah untuk menangani masalah stunting di Indonesia?
7. Menguji akurasi hasil klasifikasi.

### **1.5 Manfaat Penelitian**

Dengan penelitian ini penulis berharap dapat memberikan kontribusi kepada ilmu pengetahuan serta dapat memberikan nilai guna seperti:

#### **1. Kegunaan Teoritis**

Bagi penulis, sebagai penerapan ilmu yang penulis peroleh dan memperdalam konsep teori yang sudah dipelajari. Penelitian ini diharapkan juga dapat menjadi pengembangan ilmu pengetahuan khususnya dan di bidang Ilmu Magister Komputer umumnya.

#### **2. Kegunaan Praktis**

Dengan menganalisis sentimen publik terhadap program penurunan angka prevalensi stunting menggunakan data dari Twitter dan metode *Naïve Bayes*, akan diperoleh wawasan yang berharga tentang pandangan dan tanggapan publik terhadap program ini. Hasil analisis sentimen dapat digunakan oleh pemerintah, lembaga terkait, dan pemangku kepentingan lainnya untuk menginformasikan kebijakan dan tindakan yang lebih efektif dalam upaya pencegahan dan penurunan prevalensi stunting di Indonesia.

### **1.6 Sistematika Penulisan**

Untuk pembahasan yang lebih rinci, maka dalam penulisan penelitian ini penulis membagi atas beberapa bab, di mana satu sama lain saling berhubungan sesuai dengan ruang lingkup masalah. Secara umum gambaran isi dari masing-masing bab adalah sebagai berikut:

**BAB I            PENDAHULUAN**

Bab ini menguraikan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat penulisan, serta sistem penulisan.

**BAB II           LANDASAN TEORI**

Membahas dasar-dasar teori dan penerapan metode yang berhubungan dengan masalah-masalah yang dibahas seperti: *Stunting*, *Knowledge Discovery in Databases (KDD)*, *Data Mining*, *Sentiment Analysis*, *Information filtering*, *Text Mining*, *Machine Learning*, *Classification* dan *Naïve Bayes Classifier (NBC)*.

**BAB III          METODOLOGI PENELITIAN**

Bab ini menjelaskan tentang jenis penelitian yang dilakukan, pendekatan yang digunakan, sumber data, metoda dan alat pengumpulan data serta teknik pengolahan dan analisa data.

**BAB IV          ANALISIS DAN PERANCANGAN**

Bab ini menjelaskan prinsip kerja *Naïve Bayes Classifier (NBC)* untuk melakukan *statistical machine learning* dalam analisis sentimen untuk mengklasifikasikan data teks kedalam kelas-kelas yang telah ditentukan.

**BAB V           IMPLEMENTASI DAN HASIL**

Bab ini merupakan tahap implementasi dan penyelesaian program.

**BAB VI          PENUTUP**

Bab ini berisikan kesimpulan dan hasil pembahasan serta saran-saran.