

# BAB 1

## PENDAHULUAN

Bab 1 berisi penjelasan mengenai pendahuluan dari penelitian, dengan sub bab terdiri atas latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian yang dilakukan, dan kerangka pemikiran.

### 1.1 Latar Belakang

Wabah COVID-19 yang mulai menyebar sejak akhir tahun 2019 telah menyebabkan kematian jutaan orang di seluruh dunia. *Coronavirus Disease 2019* atau COVID-19 adalah sebutan dari penyakit yang menyerang organ pernapasan dan saraf dengan tingkat penyebaran yang sangat cepat terhadap manusia. Hingga tahun 2022, COVID-19 masih mewabah di banyak negara, termasuk Indonesia. Telah diterapkan upaya-upaya pencegahan mewabahnya COVID-19 oleh Pemerintah, seperti mewajibkan masyarakat untuk menegakkan protokol kesehatan, sosialisasi tentang COVID-19 yang terdapat di media sosial, serta turut memberikan penyediaan layanan dan pelaksanaan vaksinasi terhadap masyarakat secara menyeluruh. Kemampuan vaksin saat ini masih dalam tahap pengembangan dan belum terbukti secara kuat untuk menangkal COVID-19. Hal ini menjadi salah satu faktor yang menyebabkan keraguan dalam masyarakat perihal keperluan vaksinasi. Tidak hanya itu, keraguan dan kesalahan informasi vaksin juga menjadi hambatan besar dalam mendorong masyarakat untuk melakukan vaksinasi COVID-19.

Salah satu media yang digunakan masyarakat untuk berpendapat mengenai isu terkini yaitu melalui media sosial. Disisi lain, media sosial dapat mempengaruhi serapan informasi yang diterima seseorang. Hal yang ada pada saat ini yaitu munculnya propaganda anti vaksin yang dapat berakibat pada menurunnya kepercayaan masyarakat terhadap vaksin COVID-19. Saat ini, Twitter telah menjadi tempat untuk menyuarakan opini, pendapat, dan keluhan penggunaannya terhadap COVID-19. Media ini juga dimanfaatkan oleh Pemerintah dan lembaga kesehatan untuk mengkomunikasikan, dan mendengar suara rakyat

tentang COVID-19 yang ada di Twitter. Salah satu tolak ukur untuk menganalisis sentimen terkait isu pelayanan kesehatan dan kebijakan pemerintah saat pandemi COVID-19 dapat dilakukan dengan mengamati pembicaraan publik di Twitter. Oleh karena itu, sentimen di media sosial penting untuk diamati agar pemerintah dan lembaga kesehatan dapat menentukan strategi yang tepat dalam menangani COVID-19, khususnya dalam hal komunikasi publik.

Terdapat berbagai metode untuk melakukan klasifikasi dalam analisis sentimen, diantaranya yaitu *Support Vector Machine* (SVM), pendekatan *Lexicon*, dan *Naive Bayes Classification*. Salah satu penelitian terdahulu oleh Hayati dan Alifi (2021) mengenai analisis sentimen terkait isu vaksinasi dengan menggunakan metode SVM, menyatakan bahwa pendekatan *Lexicon* kurang efektif untuk melakukan analisis sentimen dibandingkan dengan metode SVM dan *Naive Bayes*. Hal ini dikarenakan penentuan sentimen dengan pendekatan *Lexicon* didasarkan pada keberadaan kata yang menggambarkan sentimen tertentu tanpa adanya pertimbangan terhadap domain yang dibahas secara keseluruhan dalam suatu teks. Penelitian tersebut menerapkan metode SVM dengan menggunakan tokenisasi *unigram* dan *bigram* untuk menambah akurasi analisis sentimen yang dilakukan. Dari proses evaluasi yang dilakukan, penambahan tokenisasi *bigram* pada tahap *preprocessing* terbukti dapat menambah nilai rata-rata setiap aspek pengukuran evaluasi sebesar 0,6% sampai 0,7%. Dari penelitian tersebut didapatkan nilai tertinggi seluruh aspek pengukuran dengan metode SVM mencapai 84%. Penelitian lain mengenai analisis sentimen dengan metode *Naive Bayes* salah satunya oleh Winda (2021) terkait analisis sentimen isu vaksin COVID-19 menyatakan bahwa hasil analisis yang dilakukan memiliki nilai akurasi sebesar 93%, namun tidak melakukan penggabungan *n-gram* pada proses tokenisasi.

Penerapan metode *Naive Bayes* dalam analisis sentimen memiliki keunggulan yaitu proses yang sederhana namun memiliki nilai akurasi yang tinggi. Penggabungan metode *Naive Bayes* dengan tahapan *preprocessing* terhadap *dataset* yang tepat akan menghasilkan nilai akurasi data yang lebih tinggi. Penelitian ini dilakukan dengan menerapkan metode *Naive Bayes* untuk menganalisis sentimen terkait isu vaksinasi COVID-19 dengan penggabungan

tokenisasi *unigram* dan *bigram* pada tahap *weighting dataset*. Penerapan metode ini diharapkan dapat menghasilkan proses klasifikasi yang sederhana namun memiliki nilai akurasi yang tinggi sehingga data yang dihasilkan dapat menjadi bahan atau pertimbangan dalam peningkatan pelayanan kesehatan, maupun kebijakan pemerintah dalam menangani pro kontra isu vaksinasi yang terdapat di masyarakat.

## 1.2 Perumusan Masalah

Adapun rumusan masalah pada penelitian ini yaitu sebagai berikut:

1. Bagaimana proses analisis sentimen terhadap isu vaksinasi COVID-19 dengan metode *Naive Bayes*?
2. Bagaimana pengaruh penggabungan tokenisasi *unigram* + *bigram* terhadap akurasi hasil sentimen analisis dengan metode *Naive Bayes* jika dibandingkan dengan tokenisasi *unigram* saja atau *bigram* saja?

## 1.3 Batasan Masalah

Berdasarkan rumusan masalah diatas, maka terdapat batasan masalah pada penelitian ini yaitu sebagai berikut :

1. Analisis sentimen dikhususkan terkait isu vaksinasi COVID-19, dengan menggunakan metode *Naive Bayes*.
2. *Dataset* yang digunakan yaitu berupa data *tweet* berisi isu seputar vaksinasi COVID-19 dari Twitter, dengan rentang waktu selama pengumpulan *dataset* (Mei - Oktober 2022).
3. Jumlah *dataset* yang akan digunakan yaitu 5.000 data *tweet* terbaru.
4. Pembagian data *training* dan data *testing* akan diterapkan dengan percobaan perbandingan 7:3 dan 8:2 dengan pemilihan data secara *random*.
5. Bahasa pemrograman yang digunakan adalah *Python*.

## 1.4 Tujuan Penelitian

Penelitian dilakukan atas dasar beberapa tujuan, yaitu sebagai berikut:

1. Untuk mengimplementasikan metode klasifikasi *Naive Bayes* dalam analisis sentimen.

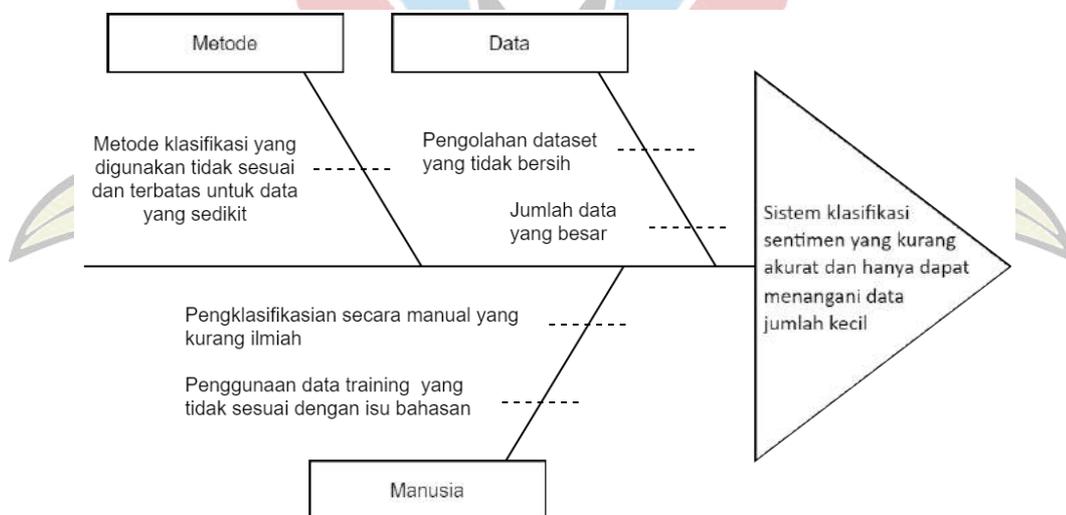
2. Untuk mengetahui pengaruh gabungan 2 jenis tokenisasi (*unigram* + *bigram*) dibandingkan dengan tokenisasi *unigram* saja atau *bigram* saja, terhadap keakuratan model klasifikasi *Naive Bayes*.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat dalam pengembangan penelitian analisis sentimen, serta diharapkan dapat menjadi dasar dalam pengambilan langkah untuk sosialisasi atau pendekatan pada masyarakat mengenai isu vaksinasi COVID-19.

### 1.6 Kerangka Pemikiran Penelitian

Kerangka pemikiran pada penelitian ini adalah sebagai berikut :



**Gambar 1.1** Kerangka Pemikiran Penelitian

Pada gambar 1.1 dapat dirumuskan beberapa faktor permasalahan penyebab dilakukannya penelitian mengenai analisis sentimen dengan metode *Naive Bayes* ini. Faktor permasalahan pertama yaitu keterbatasan kemampuan metode klasifikasi lain dalam mengatasi data dalam jumlah besar dengan akurasi yang tinggi. Pada proses analisis sentimen terdapat berbagai macam metode untuk mengklasifikasikan data. Apabila metode yang digunakan tidak sesuai dengan jenis klasifikasi, misalnya jenis data, parameter, dan sebagainya, maka hal tersebut dapat menyebabkan proses klasifikasi yang tidak maksimal. Faktor data meliputi permasalahan pada tahap *preprocessing* yang kurang efektif dalam membersihkan

data sehingga mempengaruhi akurasi analisis. Terdapat beberapa tahap *preprocessing*, dimana pemakaian dan urutan tahap tersebut harus disesuaikan dengan kebutuhan sistem untuk menghasilkan pengklasifikasian yang tepat. Metode yang digunakan dalam klasifikasi memiliki kelebihan dan kekurangan dalam memproses data dengan jumlah banyak seperti data *tweet* secara *real time*. Pengklasifikasian sentimen secara manual oleh manusia memiliki banyak kelemahan, yaitu membutuhkan banyak waktu untuk data yang banyak, dan dapat memberikan hasil yang bersifat subjektif. Pemilihan *dataset* yang akan menjadi data *training* juga harus sesuai dengan isu pembahasan, dimana dalam penelitian ini yaitu isu vaksinasi COVID-19, agar pada proses pengujian sistem dapat lebih akurat dalam pengklasifikasiannya. Faktor-faktor permasalahan tersebut tentunya dapat menjadi penyebab rendahnya tingkat akurasi analisis sentimen.

