

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada era masyarakat modern, perkembangan teknologi internet sangatlah pesat, dan media sosial telah menjadi bagian yang tidak terpisahkan. Banyak media massa yang menggunakan platform media sosial untuk menyebarkan berita. Setiap hari, berita-berita baru terus muncul dengan beragam topik. Di Indonesia, perkembangan teknologi informasi juga sangat cepat, dengan semakin bertambahnya pengguna internet. Namun, hal ini juga menyebabkan masalah persebaran berita yang sulit terkontrol, terutama berita-berita palsu (hoax).

KBBI (Kamus Besar Bahasa Indonesia) mendefinisikan berita sebagai sebuah narasi atau informasi mengenai peristiwa atau kejadian terkini. Berita ini menyajikan fakta dan juga pandangan pribadi yang menarik bagi masyarakat. Dahulu, berita disampaikan melalui media tradisional seperti surat kabar, majalah, radio, dan televisi, tetapi sekarang beralih ke platform internet.

Hoax adalah informasi palsu yang disajikan sedemikian rupa sehingga terlihat benar. Berbeda dengan rumor, ilmu semu, atau berita palsu, hoax bertujuan untuk menciptakan ketidakamanan, ketidaknyamanan, dan kebingungan di kalangan masyarakat. Dalam kondisi bingung, orang-orang dapat mengambil keputusan yang lemah, tidak meyakinkan, bahkan salah. Hoax berpotensi memberikan dampak negatif melalui tulisan dengan mempengaruhi pikiran seseorang, sementara gambar dapat menimbulkan rasa takut dan ancaman. Menurut penelitian yang dilakukan oleh Lembaga Ilmu Pengetahuan Indonesia (LIPI), masyarakat yang bersikap fanatik lebih rentan terkena hoax. Jika dibiarkan, berita hoax dapat menjadi sangat persuasif, maka diperlukan strategi untuk mengidentifikasi dan menangkal berita-berita palsu yang menyebar di media sosial. Media sosial merupakan salah satu inovasi teknologi terkini yang berperan sebagai sumber dan sarana untuk bertukar serta menerima informasi. Beberapa platform media sosial yang sudah akrab dan banyak digunakan saat ini adalah Facebook, Instagram, dan Twitter. Dalam penelitian ini, akan dilakukan analisis berita tentang

vaksin Covid-19 yang diposting di Twitter. Twitter sendiri adalah salah satu platform media sosial yang terkenal, didirikan oleh Jack Dorsey, yang digunakan untuk menyebarkan informasi berbasis mikroblogging dan mulai diluncurkan pada 13 Juli 2006.(Eka Sembodo et al., 2016).

Twitter adalah platform media sosial yang seringkali dimanfaatkan oleh individu yang tidak bertanggung jawab untuk menyebarkan berita palsu, termasuk berita hoax. Aksesibilitas Twitter sangat tinggi dan pengguna dapat dengan mudah memperoleh informasi dari pengguna lain tanpa perlu mengikuti akun tersebut terlebih dahulu. Selain itu, Twitter juga menyediakan fitur "trending topic" yang merupakan topik-topik yang sedang ramai dibicarakan di platform ini dengan menggunakan kata-kata tertentu.

Penelitian ini bertujuan untuk mengembangkan sebuah model machine learning menggunakan Naive Bayes Classifier guna mengklasifikasikan berita di Twitter. Dataset yang digunakan terdiri dari berita yang berasal dari laman resmi kominfo yang telah diberi label sebagai "hoax" dan "fakta". Setelah model terbentuk, akan digunakan untuk mengklasifikasikan berita-berita di media sosial Twitter yang belum diberi label, terkait dengan topik vaksin booster.

Naive Bayes Classifier, teknik pengklasifikasian dalam data mining, menjadi salah satu pilihan yang populer.. Sebelumnya, penelitian juga telah dilakukan dalam deteksi berita hoax di berbagai bidang. Contohnya, Hery Mustofa, dkk telah mengkaji klasifikasi berita hoax dengan menggunakan metode Naive Bayes. Hasilnya menunjukkan tingkat akurasi sebesar 85.28%, terdiri dari 307 dokumen terklasifikasi relevan dan 53 dokumen yang tidak relevan, dengan error rate sebesar 14.72%.(Mustofa & Mahfudh, 2019).

Setelah itu, hasil penelitian yang dilakukan oleh Syahnur, dkk. (2016) membahas tentang bagaimana mengkategorisasi topik dari tweet di tiga kota, yaitu Jakarta, Bandung, dan Makassar menggunakan metode multinomial naive bayes classifier. Penelitian ini menggunakan 601 dataset tweet dengan 13 kategori yang berbeda. Dalam penelitian tersebut disebutkan bahwa untuk melakukan klasifikasi kata atau tweet, diperlukan beberapa tahap sebelum proses klasifikasi dilakukan.

Salah satunya adalah tahap preprocessing, di mana tweet yang digunakan sebagai bahan pelatihan harus melalui proses tersebut. Penelitian ini berhasil mencapai nilai rata-rata f1-measure yang tinggi, yaitu sebesar 77%.

## 1.2 Rumusan Masalah

Ada pun rumusan masalah yang menjadi acuan dalam menyelesaikan skripsi ini adalah:

1. Bagaimana membangun model machine learning pengklasifikasian berita menggunakan Naïve Bayes Classifier?
2. Bagaimana menguji model machine learning pengklasifikasian berita yang belum memiliki label menggunakan Naïve Bayes Classifier?

## 1.3 Tujuan Penelitian

Ada pun tujuan dari penelitian skripsi ini adalah sebagai berikut :

1. Mengetahui cara membangun model machine learning pengklasifikasian berita dengan menggunakan metode Naive Bayes Classifier untuk klasifikasi.
2. Mengetahui hasil pengujian dari model Naive Bayes Classifier untuk digunakan dalam proses klasifikasi berita yang belum memiliki label

## 1.4 Manfaat Penelitian

Berikut merupakan beragam fungsi yang didambakan melalui penelitian ini adalah :

1. Bagi Penulis
  - a. Sebagai salah satu syarat agar penulis menyelesaikan gelar Strata satu (S1).
  - b. Hasil penelitian ini juga dapat menjadi referensi kepada penulis ketika ingin melanjutkan penelitian di masa yang akan datang.

### 2. Bagi Masyarakat

Agar masyarakat tau bagaimana cara mengklasifikasikan berita menggunakan Naïve Bayes Classifier.

### 1.5 Batasan Masalah

Pada penelitian ini terdapat beberapa poin-poin untuk membatasi permasalahan yang dapat dibahas dari penelitian ini, yaitu sebagai berikut:

1. Berita yang digunakan untuk membangun model adalah berita yang sudah diberi label hoax dan fakta dari laman resmi kominfo.
2. Berita yang digunakan untuk menguji model adalah berita yang belum mendapatkan label yang terdapat di media sosial twitter tentang vaksin booster yang di posting pada tanggal 01 Maret 2023 – 31 Mei 2023.

