

# BAB I

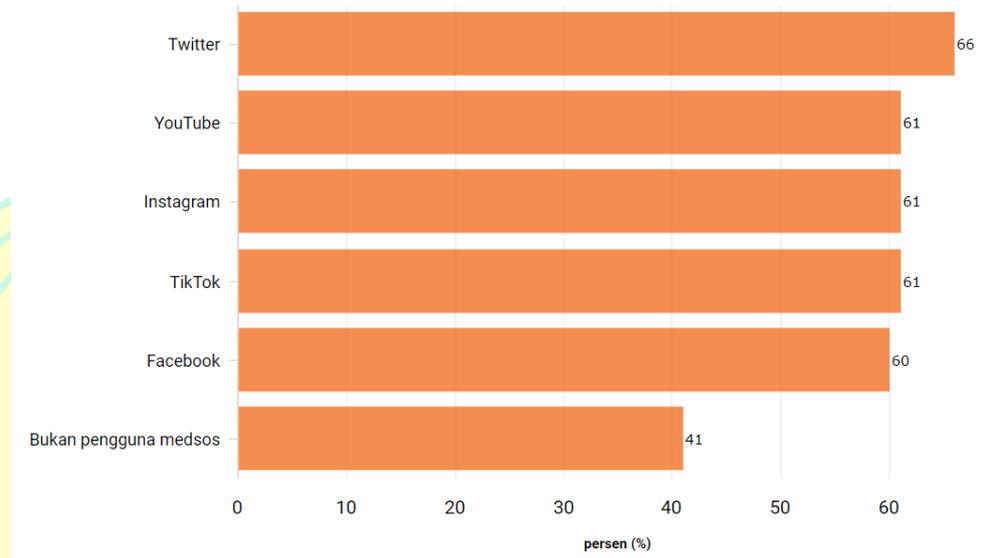
## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Hoax* merupakan liputan atau berita yang mencoba membuat pembaca tertarik mengenai kebenaran berita lalu mencoba untuk meyakinkan pembaca. Persebaran *hoax* bergantung pada siapa yang membacanya apakah akan ikut menyebarkan tanpa cek kembali kebenarannya atau dengan sengaja mengirimkan ulang kesemua orang di media sosial. Pada kamus besar bahasa Indonesia kata *hoax* dibaca menjadi hoaks, yang artinya informasi bohong. Dengan berkembangnya teknologi ini, masyarakat beralih ke media sosial sebagai alat komunikasi. Tanpa disadari *hoax* atau berita palsu juga ikut berkembang di kalangan masyarakat pengguna media sosial. Media sosial merupakan kategori wacana *online* tempat orang membuat konten dan membagikannya. Jika ada yang menyebarkan *hoax*, dapat diancam Pasal 28 ayat 1 Undang-Undang Informasi dan Transaksi Elektronik atau Undang – Undang ITE (UU ITE) yang menyatakan “Setiap orang dengan sengaja, dan tanpa hak menyebarkan berita bohong dan menyesatkan yang mengakibatkan kerugian konsumen dalam Transaksi Elektronik dapat diancam pidana berdasarkan Pasal 45A ayat (1) UU 19/2016, yaitu dipidana dengan pidana penjara paling lama 6 (enam) tahun dan/atau denda paling banyak Rp.1.000.000.000 (satu milyar rupiah).”

*Hoax* atau berita palsu terus bergulir di seluruh platform media sosial termasuk Twitter. Data Statistika (2023) dalam situs [databoks.katadata.co.id](http://databoks.katadata.co.id) menunjukkan pengguna Twitter khawatir terhadap berita asli dan palsu atau hoaks di internet dibandingkan pengguna dari media sosial lainnya. Dengan data yang di peroleh pada survei Reuters Institute yang dibukukan dalam Digital News Report 2023. Reuters Institute menyebut, proporsi yang khawatir menyentuh 66% dari total pengguna *platform* dengan nama baru X. Sementara pengguna yang khawatir dari *platform* YouTube, Instagram, hingga TikTok sebesar 61%. Selanjutnya, Facebook sebanyak 60%. Adapun orang yang tidak menggunakan media sosial tercatat sebanyak 41% yang khawatir akan berita asli dan palsu.

Reuters Institute mencatat bahwa tingkat kekhawatiran masyarakat yang lebih tinggi bukan berarti Twitter lebih banyak dibanjiri informasi palsu dibandingkan *platform* lain.



**Gambar 1. 1 Pengguna Twitter Lebih Khawatir Berita Hoaks di Internet**

(Sumber : Erlina F. Santika, 2023)

Twitter juga merupakan media sosial paling populer untuk meneliti tentang berita *hoax*. Penulis menggunakan informasi penyebaran data dari Twitter, menganalisis, dan kemudian menentukan informasi tersebut bohong atau tidak.

Teknik Word2Vec memiliki cara kerja dengan cara pemindaian kata yang terdapat pada *corpus* besar dan memiliki cara yang mirip dengan GloVe. Setelah rilis Word2Vec, metode word embeddings menjadi topik hangat pada kalangan peneliti dan diikuti dengan *word embeddings* lainnya yang muncul. GloVe adalah salah satunya. (Pennington et al., 2014). GloVe adalah model log-bilinear atau secara sederhana dapat disebut sebagai model berbasis hitungan. GloVe mempelajari hubungan kata – kata dengan menghitung seberapa sering kata – kata muncul bersama satu sama lain dalam sebuah korpus yang diberikan. Rasio probabilitas kemunculan kata – kata memiliki potensi untuk mengkodekan beberapa bentuk makna serta membantu meningkatkan kinerja pada permasalahan analogi kata. Dilihat dari risetnya, metode *word embeddings* GloVe dapat menyaingi Word2Vec pada segi akurasi, kecepatan, kemampuan

pada penyesuaian diri saat sistem. Ini yang menjadi alasan mengapa metode GloVe dipakai pada penelitian skripsi ini, untuk meminimalisir kesalahan kosakata *vocabulary*.

Berdasarkan hasil penelitian sebelumnya, belum ada yang memanfaatkan fitur ekspansi pada penerapan algoritma naïve bayes. Oleh karena itu, dalam skripsi ini, penulis bertujuan untuk mengembangkan sistem dengan menambahkan fitur ekspansi, khususnya menggunakan GloVe secara otomatis untuk mengukur tingkat kesamaan antar kata dalam korpus dan mengurangi kemiripan antar kosakata. Metode klasifikasi yang diterapkan dalam pembangunan sistem ini adalah Naïve Bayes. Tujuan dari eksperimen ini adalah menghasilkan program yang mampu mendeteksi informasi palsu (*hoax*) dengan tingkat akurasi yang tinggi, menggunakan metode klasifikasi yang telah diimplementasikan.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang yang telah dijabarkan, maka dapat diidentifikasi beberapa masalah sebagai berikut:

- a. Belum adanya penerapan algoritma Naïve Bayes pada analisis sentimen di Twitter dengan ekspansi fitur GloVe dalam pendeteksian hoaks;
- b. Mendeteksi berita hoaks pada twitter dengan fokus pada akurasi;
- c. Penggunaan GloVe untuk ekspansi fitur dalam pendeteksian hoaks untuk mengukur tingkat kesamaan antar kata dalam korpus dan mengurangi kemiripan antar kosakata.

## **1.3 Pembatasan Masalah**

- a. Penelitian ini berfokus pada studi kasus Hoaks Pemilu di Twitter;
- b. Penelitian ini menggunakan algoritma Naïve Bayes dengan ekspansi fitur GloVe untuk mencari nilai akurasi pada analisis sentimen di Twitter;
- c. Keyword yang digunakan yaitu Detikcom Pemilu, Kompas Pemilu, TurnBackHoaks;
- d. Tweet yang dianalisis hanya *tweet* Berbahasa Indonesia.

#### 1.4 Perumusan Masalah

Bagaimana penerapan algoritma Naïve Bayes dalam mengklasifikasikan berita hoaks di twitter terhadap Pemilihan Umum 2024 dan mengukur nilai akurasi dengan metode GloVe pada algoritma Naïve Bayes ?

#### 1.5 Tujuan Penelitian

Penelitian ini bertujuan untuk menganalisis hasil klasifikasi berita hoax pada twitter terhadap Pemilihan Umum 2024 menggunakan Algoritma Naïve Bayes dan mengukur nilai akurasi menggunakan ekspansi fitur GloVe dalam proses klasifikasi.

#### 1.6 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

- a. Referensi untuk menambah pengetahuan tentang klasifikasi teks dan penerapan algoritma Naïve Bayes sebagai salah satu metode untuk penelitian selanjutnya.
- b. Sebagai acuan pengembangan dan pengolahan data untuk peningkatan suatu perusahaan menggunakan algoritma Naïve Bayes dengan ekspansi fitur GloVe.

