

# BAB I

## Pendahuluan

### 1.1 Latar Belakang

Fenomena dinasti politik di Indonesia telah menjadi isu yang kontroversial dan sering memancing perdebatan publik di berbagai platform media sosial seperti YouTube dan TikTok. Isu ini berpotensi memengaruhi opini masyarakat terhadap proses demokrasi di Indonesia, sebagaimana telah diteliti dalam beberapa studi terdahulu yang menyoroti dampak dinasti politik terhadap persepsi publik (Sulaiman et al., 2021; Putra & Hakim, 2020).

YouTube dan TikTok menjadi media yang banyak digunakan untuk mengekspresikan opini terkait isu-isu politik. Berdasarkan penelitian sebelumnya, YouTube sering digunakan untuk diskusi yang lebih mendalam dan terstruktur, sedangkan TikTok lebih banyak dimanfaatkan untuk komunikasi singkat yang cenderung spontan (Kusuma et al., 2022). Perbedaan karakteristik ini menyebabkan variasi dalam pola komentar, yang berimplikasi pada analisis sentimen yang berbeda di kedua platform.

Analisis sentimen pada media sosial telah menjadi alat yang penting dalam memahami dinamika opini publik secara lebih terstruktur dan objektif (Ahmad & Rini, 2021). Dengan menggunakan algoritma pembelajaran mesin seperti Naïve Bayes dan K-Nearest Neighbors (KNN), penelitian ini bertujuan untuk membandingkan performa kedua algoritma dalam menganalisis sentimen komentar di YouTube dan TikTok, serta mengidentifikasi pola sentimen yang dominan pada kedua platform tersebut.

Penelitian ini diharapkan dapat memberikan kontribusi dalam bidang analisis opini publik serta memberikan wawasan bagi pembuat kebijakan dan akademisi dalam memahami persepsi masyarakat terkait dinasti politik di

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah yang dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana klasifikasi sentimen (positif, netral, negatif) pada komentar YouTube dan TikTok terkait isu dinasti politik di Indonesia?
2. Bagaimana proses analisis sentimen menggunakan metode Naive Bayes dan KNN untuk komentar pada kedua platform?
3. Algoritma manakah yang memiliki performa lebih baik dalam klasifikasi sentimen di antara Naive Bayes dan KNN pada dua platform yang berbeda?

## 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diidentifikasi, tujuan penelitian ini adalah:

1. Mengklasifikasikan sentimen (positif, netral, negatif) pada komentar YouTube dan TikTok terkait isu dinasti politik di Indonesia untuk memahami persepsi publik terhadap isu tersebut.
2. Menganalisis proses analisis sentimen pada komentar-komentar tersebut menggunakan metode Naive Bayes dan K-Nearest Neighbor (KNN) guna mengevaluasi mekanisme dan tahapan yang digunakan dalam penerapan kedua algoritma.
3. Membandingkan performa algoritma Naive Bayes dan KNN dalam klasifikasi sentimen untuk menentukan metode yang lebih efektif dalam mengklasifikasikan sentimen pada dataset komentar YouTube dan TikTok

## 1.4 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan untuk menjaga fokus dan ruang lingkup penelitian. Batasan ini dirumuskan sebagai berikut:

### 1. Jenis Data yang Digunakan

Penelitian ini dibatasi pada analisis komentar dari satu video di YouTube dan satu video di TikTok yang membahas isu dinasti politik. Pengambilan data dilakukan selama tiga bulan sejak tanggal publikasi video untuk menjaga konsistensi data. Selain itu, penelitian ini hanya menggunakan metode Naive Bayes dan KNN untuk klasifikasi sentimen.

### 2. Metode Klasifikasi yang Dipilih

Penelitian ini hanya menerapkan tiga metode pembelajaran mesin, yaitu Naive Bayes, K-Nearest Neighbors (KNN) Batasan ini dipilih berdasarkan popularitas metode dalam analisis sentimen teks pendek. Namun, pembatasan ini tidak mempertimbangkan algoritma lain yang mungkin memiliki performa lebih baik.

### 3. Periode Waktu Pengumpulan Data

Data komentar dikumpulkan dalam periode waktu yang terbatas, yakni saat topik terkait dinasti politik sedang menjadi perbincangan hangat. Oleh karena itu, hasil analisis mungkin tidak mencerminkan opini publik secara keseluruhan dalam jangka panjang karena adanya perubahan isu dan opini masyarakat dari waktu ke waktu.

### 4. Bahasa dan Gaya Bahasa dalam Komentar

Penelitian ini hanya fokus pada komentar dalam bahasa Indonesia. Penggunaan bahasa lokal, singkatan, atau slang yang tidak umum diabaikan dalam

proses preprocessing. Hal ini dapat menyebabkan adanya bias dalam analisis karena tidak semua variasi bahasa dipertimbangkan.

#### 5. Kategori Sentimen yang Ditetapkan

Klasifikasi sentimen dalam penelitian ini dibatasi hanya pada tiga kategori utama, yaitu positif, netral, dan negatif. Penelitian tidak mengakomodasi intensitas sentimen (misalnya sangat positif atau sangat negatif) atau kategori yang lebih spesifik yang mungkin memberikan wawasan lebih dalam mengenai opini pengguna.

#### 6. Evaluasi Performa Model

Evaluasi performa model pembelajaran mesin hanya menggunakan metrik akurasi, presisi, recall, dan f1-score. Batasan ini tidak mempertimbangkan metrik lain seperti area under the curve (AUC) yang mungkin lebih relevan untuk klasifikasi biner atau imbalanced data.

### 1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

- Bagi Akademisi dan Peneliti: Hasil penelitian ini dapat menjadi referensi bagi studi-studi selanjutnya yang ingin mengkaji analisis sentimen di media sosial, khususnya dalam konteks topik kontroversial seperti dinasti politik. Selain itu, penelitian ini memperkaya literatur mengenai efektivitas algoritma pembelajaran mesin seperti Naive Bayes dan KNN dalam klasifikasi sentimen teks pendek.
- Bagi Pembuat Kebijakan: Dengan memahami pola sentimen publik terkait dinasti politik, penelitian ini dapat memberikan masukan dalam merancang kebijakan publik yang lebih responsif terhadap kebutuhan masyarakat.

- Bagi Masyarakat Umum: Penelitian ini membantu masyarakat dalam memahami bagaimana opini publik terkait isu-isu politik diklasifikasikan berdasarkan sentimen positif, netral, dan negatif. Wawasan ini dapat meningkatkan literasi digital masyarakat, khususnya dalam menilai informasi di media sosial.

## 1.6 Sistematika Penulisan

Skripsi ini terdiri dari lima bab yang disusun sebagai berikut:

1. Bab I Pendahuluan: Bab ini menjelaskan latar belakang penelitian, rumusan masalah, tujuan, manfaat, dan sistematika penulisan.
2. Bab II Tinjauan Pustaka: Bab ini menguraikan kajian literatur terkait analisis sentimen, pembelajaran mesin, algoritma yang digunakan (Naive Bayes dan KNN), serta studi-studi terdahulu yang relevan.
3. Bab III Metodologi Penelitian: Bab ini berisi metode penelitian yang meliputi pengumpulan data, preprocessing teks, ekstraksi fitur, klasifikasi sentimen, serta evaluasi performa model.
4. Bab IV Hasil dan Pembahasan: Bab ini menyajikan hasil analisis sentimen, performa model pembelajaran mesin yang digunakan, serta pembahasan dari temuan yang diperoleh.
5. Bab V Kesimpulan dan Saran: Bab ini memberikan kesimpulan dari hasil penelitian serta saran untuk penelitian selanjutnya.