

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Analisis sentimen adalah studi komputasional yang meneliti pendapat, perasaan, dan emosi yang terkandung dalam teks. Dikenal juga sebagai *opinion mining*, bidang ini melibatkan pengolahan bahasa alami, komputasi linguistik, dan *text mining*. Tujuan utamanya adalah untuk menganalisis berbagai opini, termasuk komentar individu, yang berkaitan dengan topik, produk, layanan, organisasi, individu, atau kegiatan tertentu. Tugas utama dalam analisis sentimen adalah mengklasifikasikan polaritas yang ada dalam teks, baik dalam dokumen, kalimat, atau pendapat yang dipublikasikan secara luring maupun daring [1].

Klasifikasi adalah teknik dalam *data mining* yang memiliki fungsi dalam pengklasifikasian yang didasarkan pada sampel label kelas yang telah terklasifikasi sebelumnya. Proses ini melibatkan analisis data yang mampu menghasilkan berbagai model yang direpresentasikan oleh kelas-kelas dalam data [2]. Terdapat beragam algoritma klasifikasi yang menerapkan prinsip statistika dan probabilitas, salah satunya yaitu algoritma *Naive Bayes*. Terdapat berbagai algoritma klasifikasi yang menggunakan prinsip statistik dan probabilitas, salah satunya adalah algoritma *Naive Bayes*. Algoritma ini digunakan untuk melakukan klasifikasi berdasarkan data historis. *Naive* mengacu pada asumsi bahwa semua atribut dalam data bersifat independen satu sama lain [3]. Seperti pada [4][5] penelitian yang menggunakan sumber data dari media sosial *Twitter* dengan menerapkan algoritma *Naive Bayes* untuk opini feminisme memperoleh akurasi sebesar 87%, dan untuk persepsi masyarakat Indonesia terhadap Pemilu yang diselenggarakan pada tahun 2019 memperoleh akurasi sebesar 81%. Selain itu, penelitian [6] melibatkan algoritma *Naive Bayes* untuk analisis sentimen berkaitan dengan belajar *programming* pada media sosial *YouTube* dengan perolehan akurasi sebesar 80,99%.

Selanjutnya, ada algoritma *Logistic Regression* yang juga menggunakan prinsip probabilitas dan statistika. Algoritma ini bersifat matematis dan bertujuan untuk mengidentifikasi hubungan antara satu variabel dependen biner dengan satu

atau lebih variabel independen. Variabel independen ini dapat berbentuk biner, diskrip, kontinu, atau kombinasi dari semuanya. *Logistic Regression* membutuhkan jumlah sampel atau data yang besar untuk menghasilkan hasil yang akurat [7]. Dalam penelitian [8][9] analisis sentimen dengan menggunakan algoritma *Logistic Regression* pada *platform online* untuk judul berita pada Pemilu 2019 mendapatkan nilai akurasi sebesar 86%, dan untuk penerbangan maskapai *Lion Air* memperoleh hasil akurasi sebesar 82%. Selanjutnya pada penelitian mengenai analisis sentimen kendaraan listrik yang didasarkan pada data media sosial *Twitter* dengan menggunakan algoritma *Logistic Regression* mendapatkan perolehan nilai persentase sejumlah 86% [10].

Pemilu 2024, khususnya pemilihan umum untuk Calon Presiden Dan Wakil Presiden, diatur oleh Komisi Pemilihan Umum (KPU) sesuai dengan peraturan KPU Nomor 15 Tahun 2023. Aturan ini mencakup pelaksanaan debat yang diadakan sebanyak tiga kali untuk Calon Presiden dan dua kali untuk Calon Wakil Presiden. Debat-debat ini disiarkan melalui TV swasta maupun TV negara, dan rekamannya kemudian diunggah ke *platform* media sosial *YouTube*. *YouTube* menyediakan kolom komentar untuk para penonton, terlebih khusus masyarakat Indonesia agar dapat menuangkan opini mereka terkait hasil unggahan debat pasangan calon yang ditonton. Opini publik berperan penting terhadap elektabilitas para pasangan calon. Klasifikasi terhadap analisis sentimen melahirkan pelacakan dan analisis dari pola, opini, dan komentar yang muncul dalam berbagai *platform* media sosial, terlebih khusus pada *platform* kolom komentar yang terdapat pada *YouTube* [11].

Berdasarkan uraian masalah di atas, klasifikasi analisis sentimen debat calon presiden dan calon wakil presiden, akan dilakukan perbandingan algoritma *Naive Bayes* dan *Logistic Regression*. Pemilihan kedua algoritma ini didasarkan pada perolehan hasil akurasi dari penelitian-penelitian terkait yang bernilaiimbang dan menggunakan prinsip probabilitas dan statistika. Perbandingan kedua algoritma digunakan untuk mencari efektivitas akurasi, *precision*, dan *recall* terhadap sentimen masyarakat Indonesia yang berhubungan dengan debat calon presiden dan calon wakil presiden Indonesia yang diambil pada komentar *Youtube* dari siaran KompasTV, Metro, TVRI, SCTV, dan TvOne.

## 1.2 Rumusan Masalah

Bagaimana membangun aplikasi yang dapat membandingkan algoritma *Naïve Bayes* dan *Logistic Regression* dengan studi kasus debat Calon Presiden dan Calon Wakil Presiden Republik Indonesia Tahun 2024-2029?

## 1.3 Tujuan Tugas Akhir

Membangun aplikasi yang dapat membandingkan algoritma *Naïve Bayes* dan *Logistic Regression* dengan studi kasus debat Calon Presiden dan Calon Wakil Presiden Republik Indonesia Tahun 2024-2029, sehingga dapat menentukan algoritma yang lebih efektif dalam menganalisis dan mengklasifikasi hasil debat tersebut, dan memberikan wawasan lebih dalam tentang kinerja masing-masing algoritma dalam konteks pemilu.

## 1.4 Manfaat Tugas Akhir

1. Mengevaluasi, dan membandingkan efektivitas algoritma *Naïve Bayes* dan *Logistic Regression* dalam menganalisis data debat Calon Presiden dan Calon Wakil Presiden.
2. Membantu dalam memahami kinerja masing-masing algoritma, dan memberikan wawasan yang lebih baik dalam sentimen masyarakat Indonesia yang berhubungan dengan debat para Calon Presiden dan Calon Wakil Presiden pada Pemilu 2024.
3. Membantu pengambil keputusan dan peneliti dalam memilih metode yang paling tepat untuk aplikasi serupa di masa depan.
4. Membantu para jurnalistik dalam memperoleh data terkait sentimen masyarakat Indonesia.

## 1.5 Batasan Masalah

Berikut dipaparkan batasan-batasan terkait Tugas Akhir:

1. Menggunakan sumber data komentar berbasis bahasa Indonesia yang berasal dari media sosial *Youtube*, yaitu 5 jumlah debat yang meliputi 3 debat calon presiden dan 2 debat calon wakil presiden.
2. Data berasal dari video debat calon presiden dan calon wakil presiden

Indonesia pada *channel Youtube* KompasTV, Metro, TVRI, SCTV, dan TvOne.

## 1.6 Metode Penelitian

1. Identifikasi Masalah; Pada tahap ini dilakukan proses yang bertujuan dalam memahami masalah yang termuat dari berlangsungnya penelitian, yang diawali dengan identifikasi masalah utama untuk menjadi penentu dalam pengangkatan alur penelitian.
2. Penentuan Tujuan; Tahapan ini diperuntukkan sebagai penentuan tujuan yang didasarkan dari hasil identifikasi masalah pada tahap sebelumnya untuk menjadi penentu tujuan penelitian yang akan dilakukan.
3. Studi Literatur; Tahap ini merupakan pendekatan pencarian informasi yang memiliki relevansi dengan penelitian, serta teori yang menjadi indikator pendukung penelitian.
4. Analisis dan Pengumpulan Data merupakan kegiatan menganalisis terkait permasalahan yang ditemukan beserta dengan pengumpulan beragam data yang dibutuhkan terhadap penelitian yang diangkat dari *YouTube* pada tanggal 25 Februari 2024.
5. Perancangan; Dalam tahap ini dilakukan pembuatan *storyboard*, diagram pemodelan, beserta dengan algoritma *Naive Bayes* dan *Logistic Regression*.
6. Implementasi merupakan hasil implementasi ke dalam bentuk pengodean dalam bahasa pemrograman *Python* yang disesuaikan dengan hasil buatan rancangan.
7. Pengujian dilakukan terhadap hasil implementasi pengodean algoritma, dengan melihat hasil dari performa tingkat akurasi dari data uji.

## 1.7 Sistematika Penulisan

Tugas Akhir ini tersusun dalam beberapa struktur bagian beserta pembahasan sebagai berikut:

Bab I Pendahuluan: Bab ini memuat tentang latar belakang, rumusan masalah, tujuan Tugas Akhir, manfaat Tugas Akhir, batasan masalah, metodologi penelitian, dan sistematika penulisan.

Bab II Studi Pustaka: Bab ini berisi tentang teori-teori pendukung yang

memiliki kaitan dengan topik penelitian yang diangkat, penelitian-penelitian sebelumnya yang berhubungan atau serupa yang dijadikan sebagai referensi untuk penulis dalam melakukan pengerjaan Tugas Akhir.

Bab III Analisis: Dalam Bab ini mengandung identifikasi masalah, analisis kebutuhan, analisis data, beserta dengan perhitungan dari *Naive Bayes* dan *Logistic Regression*.

Bab IV Perancangan: Bab ini memuat tentang rencana perancangan pembangunan aplikasi analisis sentimen yang berkaitan dengan pemodelan diagram serta rancangan tampilan antarmuka pengguna.

Bab V Implementasi: Bab ini membahas terkait tahapan-tahapan yang berkaitan dengan pembuatan aplikasi beserta dengan integrasi atau implementasi dari penggunaan algoritma yang berkaitan dengan persyaratan aplikasi, implementasi antarmuka pengguna, implementasi basis data aplikasi, dan implementasi pengodean program.

Bab VI Pengujian: Bab ini mengandung tentang hasil uji terhadap implementasi pengodean yang dilakukan sebelumnya berdasarkan studi kasus pengujian.

Bab VII Kesimpulan dan Saran: Bab terakhir ini berisikan kesimpulan terhadap berhasilnya Tugas Akhir yang dikerjakan, beserta saran yang bertujuan untuk pengembangan di kemudian hari.

