

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen adalah metode pemrosesan bahasa alami yang digunakan untuk mengidentifikasi dan mengukur perasaan atau pendapat yang terkandung dalam teks. Banyak penelitian telah dilakukan untuk mengembangkan metode analisis sentimen, termasuk metode *Support Vector Machine* (SVM), pendekatan berbasis kosakata, dan arsitektur analisis sentimen berbasis *deep learning* (Fatihin, Susanto, & Fetrina, 2021).

Pada masa kini, media sosial sudah menjadi kebutuhan dalam kehidupan sehari-hari, digunakan untuk berbagai tujuan, termasuk untuk keperluan pribadi dan bisnis. Media sosial populer seperti *Facebook*, *Twitter*, *Instagram*, dan *WhatsApp* memberi orang *platform* untuk terhubung dan berbagi pandangan. Dengan adanya berbagai jenis konten di jejaring sosial, seperti komentar, gambar, *emoticon*, video, dan lainnya, orang dapat dengan bebas mengungkapkan pendapat mereka (Permatasari, Linawati, & Jasa, 2021).

Perdebatan politik akan menjadi fokus utama pemberitaan dan media sosial saat pemilihan umum (pemilu) 2024 semakin dekat. Pengguna media sosial bebas mendapatkan informasi, berbagi pendapat, dan mengkritik. Namun, karena kebebasan berekspresi yang tinggi di media sosial, pengguna dapat membuat konten yang mengandung ujaran kebencian tanpa disadari. Fenomena ini dapat menimbulkan masalah besar karena UU ITE mengatur ujaran kebencian (Setyani, 2022). Penelitian sebelumnya oleh Muftia Calida dan M. Didik R. menggunakan algoritma *Naïve Bayes* untuk menganalisis sentimen ujaran kebencian pada pemilihan presiden 2019. Penelitian ini menggunakan kata kunci pilpres 2019 di beberapa kota di Indonesia dan mengumpulkan 5055 data. Hasil klasifikasi menunjukkan sentimen *irrelevant* sebesar 11,3% dengan 573 data, sentimen negatif sebesar 35,4% dengan 1786 data, sentimen netral sebesar 26,7% dengan 1350 data, dan sentimen positif sebesar 26,6% dengan 1343 data. Di antara lima kota tersebut, sentimen negatif mendapat skor tertinggi, dengan nilai 35,4% (Chalida & R. Wahyudi, 2020).

Ujaran kebencian adalah definisi kejahatan yang melibatkan makian atau penghinaan terhadap orang atau kelompok berdasarkan ras, jenis kelamin, orientasi seksual, suku atau agama. Ini adalah jenis penghinaan yang menimbulkan permusuhan dan intimidasi serta menyebabkan pencemaran lingkungan (Rohani, 2018). Di media sosial, ujaran kebencian dapat berupa pesan yang menghina, mendiskriminasi, atau mengancam individu atau kelompok berdasarkan atribut seperti suku, agama, ras, atau bahkan pandangan politik. Analisis Jenis dan Makna Pragmatis Ujaran Kebencian di Media Sosial Twitter Tahun 2023 adalah penelitian yang dilakukan oleh K.N. Widyatnyana et al. Penelitian ini bertujuan untuk meningkatkan pemahaman masyarakat tentang ujaran kebencian yang muncul di media sosial. Penemuan-penemuan ini merupakan solusi untuk membantu masyarakat memahami jenis-jenis ujaran kebencian ini, yang mencakup penghinaan, pencemaran nama baik, penodaan agama, gangguan, berita palsu, dan provokasi atau hasutan. Namun, makna pragmatis seperti sarkasme, menggambarkan seorang pemimpin, menyombongkan diri, mempertanyakan, kekecewaan, dan ajakan ditemukan (Widyatnyana, Rasna, & Putrayasa, 2023). Penyebaran ujaran kebencian dapat mengancam kredibilitas pemilu, mengganggu diskusi publik, dan bahkan dapat menyebabkan ketegangan sosial. Akibatnya, diperlukan suatu sistem yang dapat mengidentifikasi ujaran kebencian dalam kalimat.

Untuk masalah klasifikasi dan regresi, Support Vector Machine (SVM) mencari *hyperplane* atau fungsi pemisahan terbaik untuk memisahkan kelas-kelas dalam ruang berdimensi N , di mana N adalah jumlah fitur dalam data. Dalam penelitian tahun 2021 Primadani A. dan Retno W., Analisis Sentimen Wacana Pemandangan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine* (SVM).

Hasil pengujian menggunakan SVM terhadap 1.236 *tweet* (404 positif dan 832 negatif) menunjukkan bahwa SVM lebih baik dari metode sebelumnya (BM25 + KNN dan *Naïve Bayes*). Dengan akurasi 96,68%, *precision* 95,82%, *recall* 94,04%, dan AUC 0,979, Salah satu keunggulan utama *Support Vector Machine* adalah kemampuannya untuk menyelesaikan masalah yang berkaitan dengan sampel data yang sangat besar dan luas. Karena kompleksitas komputasinya yang rendah dibandingkan dengan banyak algoritma lainnya, metode ini banyak digunakan. Salah satu kelebihan lain dari penggunaan *Support Vector Machine* adalah penentuan

jarak, yang memungkinkan penggunaan *Support Vector* untuk membedakan dua kelompok data dari dua kelas yang berbeda (Rohman, Berlilana, & Arsi, 2021).

Dalam penelitian ini, sistem berbasis web dikembangkan untuk mengidentifikasi ujaran kebencian dalam kalimat bahasa Indonesia. *Support Vector Machine* (SVM) digunakan untuk melakukan klasifikasi sentimen di situs web ini. Data yang digunakan untuk melatih dan memeriksa sistem berasal dari komentar dan *tweet* yang berkaitan dengan pemilu 2024.

1.2 Rumusan Masalah

Untuk mengidentifikasi masalah tersebut, sebuah rumusan masalah dibuat berdasarkan latar belakang di atas:

1. Bagaimana implementasi metode *support vector machine* sebagai model untuk mengklasifikasikan teks yang termasuk ujaran kebencian sehingga dapat digunakan untuk mendeteksi teks yang mengandung unsur ujaran kebencian?
2. Bagaimana tingkat akurasi metode *support vector machine* dalam mengklasifikasikan teks yang termasuk ujaran kebencian sehingga dapat digunakan untuk mendeteksi teks yang mengandung unsur ujaran kebencian?

1.3 Batasan Masalah

Batasan masalah yang diperlukan untuk menyelesaikan penelitian ini adalah sebagai berikut:

1. Penelitian ini akan berfokus pada *tweet* di *twitter* atau komentar pada postingan dari media sosial *facebook* dan *instagram* yang memiliki keterkaitan dengan pemilu tahun 2024.
2. Data teks yang digunakan pada penelitian ini hanya teks yang berbahasa Indonesia.
3. Pada penelitian ini hanya memakai model *support vector machine* sebagai model klasifikasi utama nya.
4. Pada penelitian ini data yang dapat dipakai untuk *training* hanya kumpulan dataset yang sudah diberi label positif negatif.
5. Pada penelitian ini *website* yang dikembangkan hanya bisa melakukan proses klasifikasi dari file *excel* berupa *csv* yang berisi data teks hasil *scraping* dari media sosial.

6. Dalam penelitian ini, tidak ada bahasa pemrograman lain selain *python*, yang digunakan dengan *framework flask 2.3*.
7. *Database* yang digunakan untuk membangun *website* pada penelitian ini adalah MySQL.

1.4 Tujuan

Berikut adalah tujuan dari penelitian ini:

1. Mengimplementasikan metode *support vector machine* untuk proses pengklasifikasian data dengan menggunakan data hasil *scraping* terkait pemilu tahun 2024 pada media sosial yaitu *facebook, instagram, dan twitter*.
2. Mengetahui tingkat akurasi metode *support vector machine* dalam mengklasifikasikan teks yang termasuk ujaran kebencian sehingga dapat digunakan untuk mendeteksi teks yang mengandung unsur ujaran kebencian.

1.5 Manfaat

Adapun manfaat yang diharapkan dalam penelitian ini adalah sebagai berikut:

1. Sistem yang dibuat dapat mengklasifikasikan sebuah teks termasuk ujaran kebencian atau bukan.
2. Memberikan hasil klasifikasi dari metode *support vector machine* terhadap komentar maupun sebuah *tweet* dari media sosial.
3. Sistem dapat dijadikan referensi untuk penelitian selanjutnya yang juga akan membahas terkait analisis sentimen menggunakan *support vector machine*.

1.6 Sistematika Penulisan

Agar mempermudah pemahaman pada pembahasan penulisan skripsi ini, maka sistematika penulisan diperoleh sebagai berikut:

- BAB I : Dalam pendahuluan akan menemukan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan, keuntungan, metode penelitian, dan sistematika penelitian.
- BAB II : Tinjauan pustaka mencakup dasar teori penelitian.
- BAB III : Analisis dan Perancangan Sistem mencakup perancangan sistem yang menggunakan diagram flow dan desain struktur menu.
- BAB IV : Konsep yang diusulkan dibahas secara rinci dan detail dalam Implementasi dan Pengujian, yang mencakup desain sistem yang akan menggunakan konsep tersebut.

BAB V : Penutup mengandung kesimpulan dan rekomendasi untuk penelitian lanjutan.