

BAB I

LATAR BELAKANG

1.1 Latar Belakang

Platform digital telah menjadi ruang utama bagi masyarakat Indonesia dalam menyampaikan opini, kritik, dan respons terhadap isu-isu publik secara cepat dan masif [1]. Platform seperti Twitter atau X memberikan kesempatan kepada masyarakat untuk menyebarluaskan pendapat politik secara cepat dan dalam jumlah yang banyak, menciptakan data teks yang tidak teratur yang menggambarkan perubahan opini publik secara langsung. [2]. Salah satu isu yang menimbulkan respons luas di media sosial adalah kebijakan kenaikan gaji anggota Dewan Perwakilan Rakyat (DPR), yang memicu perdebatan mengenai keadilan sosial, transparansi dalam perumusan kebijakan, serta penentuan prioritas anggaran negara [3]. Besarnya volume komentar, keragaman gaya bahasa, serta munculnya unsur linguistik seperti ironi dan sarkasme menyebabkan analisis manual menjadi tidak efektif, sehingga diperlukan pendekatan komputasi berbasis *machine learning* untuk mengolah dan mengklasifikasikan sentimen masyarakat melalui pendekatan yang terencana dan kuantitatif [4].

Dalam konteks pengolahan data teks digital, analisis sentimen berperan penting dalam mengidentifikasi orientasi opini publik dengan mengonversi data tidak terstruktur menjadi informasi yang terorganisasi [5]. Studi-studi yang sudah ada memaparkan penerapan pendekatan *machine learning* bekerja dengan baik pada klasifikasi sentimen di data media sosial berbahasa Indonesia. Di antara algoritma yang umum digunakan, *Naive Bayes* menonjol karena pendekatannya yang sederhana, efisiensi dalam pengolahan data berukuran besar, dan performanya yang cukup stabil pada tugas klasifikasi teks [6]. *Naive Bayes* juga sering diterapkan pada isu politik di media sosial karena kestabilannya dalam mengklasifikasikan sentimen positif, negatif, dan netral [2]. Namun demikian, karakteristik bahasa media sosial yang tidak baku, mengandung sarkasme, serta memiliki makna kontekstual yang kompleks dapat menjadi kendala bagi model probabilistik seperti *Naive Bayes* dalam menghasilkan klasifikasi yang presisi [7].

Tidak seperti pendekatan lainnya, Support Vector Machine atau (SVM) kerap dipilih dalam analisis sentimen karena kemampuannya menentukan garis pemisah terbaik antar kategori dengan memaksimalkan jarak antar kelompok, serta tetap bekerja secara optimal meskipun jumlah fitur yang digunakan sangat besar [1]. *SVM* mampu menangani pola data yang tidak linear serta variasi linguistik yang umum ditemukan dalam percakapan media sosial [8]. Sejumlah studi melaporkan bahwa *Support Vector Machine* (SVM) cenderung menghasilkan tingkat akurasi yang lebih tinggi dibandingkan *Naive Bayes* dalam tugas klasifikasi sentimen, baik pada ulasan aplikasi, opini konsumen, maupun pembahasan isu sosial dan politik [9]. Pada situasi ketidakseimbangan distribusi data, *Support Vector Machine* (SVM) menunjukkan kinerja yang lebih stabil dan cenderung memberikan hasil prediksi yang lebih akurat [10]. Perbedaan karakteristik ini menunjukkan bahwa *SVM* lebih adaptif dalam menangani kompleksitas bahasa, sedangkan *Naive Bayes* tetap unggul dari sisi kecepatan dan efisiensi komputasi.

Berdasarkan perbedaan karakteristik tersebut, penelitian ini menjadikan *Naive Bayes* dan *Support Vector Machine* (SVM) sebagai dua metode yang relevan untuk dibandingkan dalam analisis sentimen publik terhadap kebijakan kenaikan gaji DPR. Karakteristik data opini publik yang berjumlah besar, bersifat tidak terstruktur, serta memiliki keragaman bahasa yang tinggi menuntut penerapan metode yang tidak hanya efisien, tetapi juga mampu menghasilkan tingkat akurasi yang baik [6]. Oleh karena itu, penelitian ini memfokuskan pada perbandingan kinerja kedua algoritma dalam mengklasifikasikan sentimen positif, netral, dan negatif dengan menggunakan metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score* [7]. Hasil perbandingan tersebut diharapkan mampu memberikan gambaran yang lebih objektif terkait keunggulan dan keterbatasan masing-masing metode dalam penerapan analisis sentimen [9], serta berperan sebagai rujukan dalam pengambilan keputusan berbasis data untuk mendukung evaluasi kebijakan publik di Indonesia [5].

1.2 Rumusan Masalah

Berdasarkan informasi yang telah disampaikan, pertanyaan inti dari penelitian ini adalah:

1. Bagaimana cara menggunakan metode SVM dan Naive Bayes untuk menganalisis sentimen opini publik tentang kebijakan kenaikan gaji DPR di media sosial?
2. Bagaimana cara menguji hasil analisis dari kedua metode tersebut untuk mengetahui seberapa akurat dalam membedakan sentimen positif, netral, dan negatif?
3. Antara SVM dan Naive Bayes, metode mana yang memberikan hasil paling baik dengan mengacu pada metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Tujuan

Mengacu pada rumusan masalah yang telah disusun, tujuan penelitian ini dirumuskan sebagai berikut:

1. Mengklasifikasikan sentimen opini publik di media sosial mengenai kebijakan kenaikan gaji DPR menjadi tiga kelas sentimen yang terdiri atas positif, netral, dan negatif.
2. Membandingkan output analisis sentimen yang dihasilkan oleh metode SVM dan Naive Bayes.
3. Menilai metode mana yang memiliki kinerja terbaik dengan mengacu pada metrik evaluasi yang meliputi *accuracy*, *precision*, *recall*, dan *F1-score*.

1.4 Batasan Masalah

Agar ruang lingkup penelitian tetap terkontrol dan sesuai dengan tujuan yang ditetapkan, maka batasan masalah dalam penelitian ini dirumuskan sebagai berikut:

1. Penelitian berfokus pada data opini publik di media sosial yang berkaitan dengan isu kebijakan kenaikan gaji DPR.
2. Ruang lingkup data penelitian ini terbatas pada data teks berbahasa Indonesia.

3. Adapun model klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) dan *Naive Bayes*, dengan fokus pada analisis perbandingan kinerja kedua metode tersebut.
4. Dataset yang digunakan sudah melalui proses pelabelan menjadi tiga kelas sentimen: positif, netral, dan negatif.
5. Website yang dikembangkan hanya dapat memproses file CSV berisi data teks hasil scraping dari media sosial untuk dianalisis dan diklasifikasikan.

1.5 Manfaat

Adapun manfaat yang diharapkan dari pelaksanaan penelitian ini adalah sebagai berikut:

1. Sistem dapat mengklasifikasikan opini publik di media sosial menjadi sentimen positif, netral, atau negatif terkait kebijakan kenaikan gaji DPR.
2. Penelitian ini menghasilkan perbandingan kinerja antara penerapan metode *Support Vector Machine* (SVM) dan *Naive Bayes* untuk analisis opini publik pada media sosial.
3. Sistem yang dikembangkan dapat dijadikan sebagai referensi dan dasar pengembangan bagi penelitian selanjutnya membahas analisis sentimen kebijakan publik menggunakan algoritma klasifikasi.

1.6 Sistematika Penulisan

Sebagai upaya untuk memudahkan pemahaman terhadap isi tugas akhir, sistematika penulisan disusun secara sistematis di bawah ini:

BAB I : Latar belakang riset ini berangkat dari diskusi mengenai kenaikan gaji anggota dewan. Permasalahan dirumuskan melalui pengamatan terhadap isu yang sedang berkembang, disertai peninjauan data lapangan. Fokus penelitian tidak hanya pada penilaian opini warganet, tetapi juga pada perbandingan dua teknik pendeteksian sentimen. Namun demikian, ruang lingkup kajian dibatasi pada cuitan Twitter dalam periode tertentu. Manfaat penelitian tercermin dari kontribusinya dalam memperkaya pemahaman mengenai dinamika opini di ranah digital. Penyajian pembahasan disusun secara bertahap, dimulai dari landasan teori hingga pemaparan hasil akhir.

Sorotan utama terletak pada perbedaan perilaku SVM dibandingkan dengan Naive Bayes dalam menginterpretasikan emosi teks. Pendekatan ini dipilih karena kemampuannya mengurai narasi yang kompleks tanpa menghilangkan makna yang terkandung.

- BAB II : Membahas teori dan penelitian yang mendasari penelitian ini, seperti konsep media sosial, opini publik, analisis sentimen, NLP, Naive Bayes, dan SVM.
- BAB III : Pengumpulan data menjadi langkah awal dalam bab ini, kemudian dilanjutkan dengan proses *preprocessing* teks agar data lebih bersih sebelum digunakan. Metode seperti *Support Vector Machine* diterapkan setelah tahap persiapan data selesai, sementara *Naive Bayes* juga digunakan sebagai pembanding hasil. Evaluasi tidak hanya berfokus pada nilai akurasi, tetapi setiap model turut dinilai berdasarkan presisi, kemampuan dalam menangkap kategori yang benar, serta skor *F1* yang menggabungkan kedua aspek tersebut.
- BAB IV : Pada bagian ini disajikan hasil pengujian model, analisis perbandingan kinerja antar metode klasifikasi yang digunakan, serta pembahasan mengenai hasil klasifikasi sentimen yang diperoleh.
- BAB V : Bab ini berisi rangkuman temuan penelitian serta saran yang dapat dijadikan landasan bagi pengembangan penelitian pada waktu yang akan datang.