

BAB I

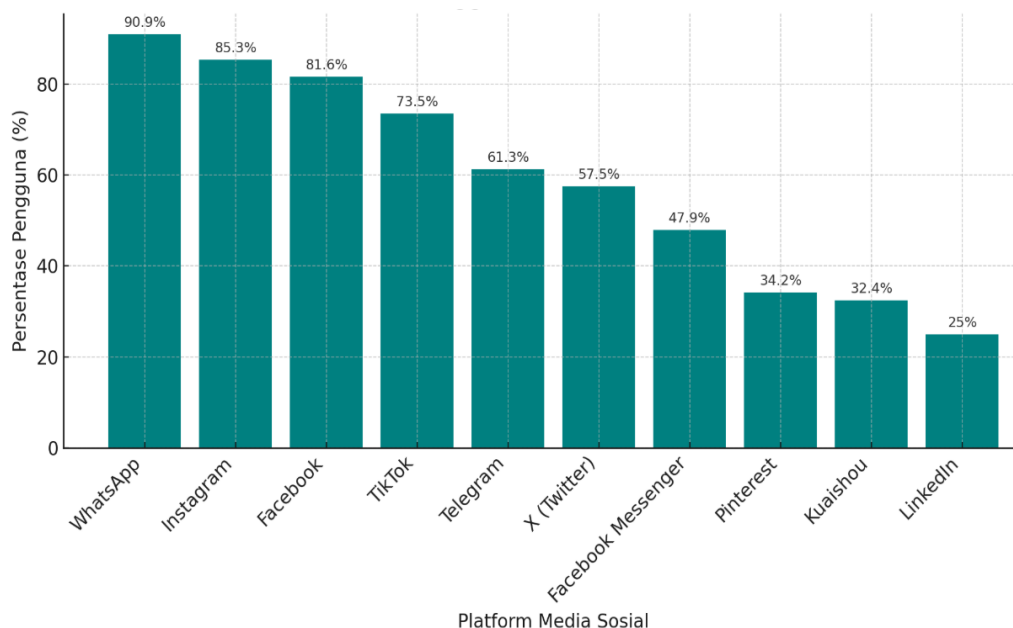
PENDAHULUAN

1.1 Latar Belakang

Pemilihan Umum (Pemilu) yang diselenggarakan pada tanggal 14 Februari 2024 di Indonesia menarik perhatian publik karena menimbulkan berbagai dinamika politik. Achmad (2024) menyatakan bahwa Pemilu kali ini merupakan salah satu yang paling problematik sejak era reformasi. Badan Pengawas Pemilihan Umum (Bawaslu) menerangkan hingga 8 Januari 2024 atau 36 hari menjelang pemungutan suara, terdapat sejumlah 1.032 dugaan pelanggaran. Berbagai dugaan pelanggaran tersebut berpotensi memperburuk citra pemerintah sehingga menurunkan tingkat kepercayaan masyarakat terhadap pemimpin yang terpilih.

Tingginya partisipasi masyarakat dalam mengawasi proses Pemilu tidak terbatas pada periode sebelum pemungutan suara, melainkan berlanjut hingga pasca Pemilu. Meskipun proses pemungutan suara telah selesai, masyarakat tetap aktif mengawasi dan menyuarakan opini melalui media sosial. Media sosial menjadi wadah utama bagi masyarakat dalam mengekspresikan pandangan terhadap jalannya Pemilu maupun tokoh politik. Menurut laporan *We Are Social* pada tahun 2024, penggunaan media sosial di Indonesia menempati rata-rata durasi terlama kesembilan dari seluruh pengguna internet global. Masyarakat tertarik dengan kemudahan media sosial untuk berbagi informasi secara cepat dan tanpa batasan. Sebagai alat komunikasi yang populer dan banyak digunakan, media sosial memiliki peran penting di negara demokrasi. Salah satu media sosial yang menjadi

sumber berita secara *real time* mengenai Pemilu 2024 adalah X, yang sebelumnya dikenal sebagai *Twitter*.



Gambar 1. Persentase Pengguna Media Sosial di Indonesia

Sumber: *We Are Social* (2024)

Berdasarkan laporan *We Are Social* pada Gambar 1, media sosial X menempati peringkat keenam dari sepuluh media sosial yang terbanyak digunakan di Indonesia. Peningkatan pengguna X dari tahun ke tahun membuat jumlah opini yang ditulis oleh masyarakat menjadi semakin masif. Opini yang disampaikan melalui postingan X menyimpan informasi yang bernilai untuk dianalisis. Respon masyarakat dalam media sosial dapat mencerminkan tingkat kepuasan, harapan, maupun kritik terhadap pasangan presiden dan wakil presiden terpilih.

Metode yang umum digunakan untuk mengolah respon masyarakat menjadi bahan evaluasi yang baik untuk disajikan adalah dengan *text mining*. *Text mining* mengekstraksi kata kunci, pendapat, dan ulasan teks sehingga dapat mendukung dalam memahami pendapat masyarakat (Kurniawan dan Susanto, 2019). Salah satu

penerapan dari *text mining* untuk data di media sosial adalah dengan analisis sentimen. Analisis sentimen adalah suatu proses pengolahan data yang berbentuk teks dan bertujuan untuk memahami, mengekstraksi, dan mengolah informasi yang terkandung dalam rumor yang kontroversial (Puad *et al.*, 2023).

Dalam bidang klasifikasi sentimen, metode *Support Vector Machine* (SVM) telah terbukti memberikan hasil yang unggul. Penelitian Guia *et al.* (2019) membandingkan empat metode klasifikasi yaitu *Naïve Bayes*, SVM, *Decision Tree*, dan *Random Forest* menggunakan 400.000 data ulasan ponsel pada web Amazon.com. Hasilnya menunjukkan bahwa SVM memiliki akurasi tertinggi yaitu 89%, diikuti oleh metode *Random Forest* sebesar 88%, metode *Naïve Bayes* 85%, dan terakhir metode *Decision Tree* 82%. Pada penelitian tersebut, metode SVM dianggap sebagai metode yang komplik karena tidak hanya menghasilkan nilai akurasi yang tinggi, tetapi juga menunjukkan performa yang baik pada metrik evaluasi lainnya seperti *precision*, *recall*, dan *F1-score*.

Pada dasarnya, metode SVM dirancang untuk mengklasifikasikan data ke dalam dua kelas (*binary classification*). Dalam konteks analisis sentimen, data sering diklasifikasikan ke dalam tiga kelas yaitu negatif, netral, dan positif. Perlu digunakan pendekatan tambahan agar SVM dapat mengatasi permasalahan klasifikasi *multiclass*. Dua pendekatan yang umum digunakan adalah *One Versus One* (OVO) dan *One Versus Rest* (OVR). Alita *et al.* (2020) telah melakukan penelitian dengan 2000 data opini publik dengan label sentimen tiga kelas. Penelitian tersebut membandingkan SVM-OVO dan SVM-OVR. Hasilnya menunjukkan bahwa metode SVM dengan pendekatan OVR memiliki nilai akurasi yang lebih tinggi dengan selisih sebesar 0,06.

Teknik pembobotan kata pada penelitian ini menggunakan *Term Frequency – Inverse Document Frequency* (TF-IDF) yang dikombinasikan dengan ekstraksi fitur berbasis *N-Gram* guna meningkatkan pemahaman konteks kata. *N-Gram* memecah teks menjadi urutan kata secara berpasangan agar makna frasa dalam teks tetap terjaga, terutama dalam konteks kalimat negasi yang seringkali kehilangan makna jika diproses sebagai kata tunggal. Penelitian yang dilakukan oleh Anjani dan Fauzan (2021) membuktikan bahwa penggunaan *N-Gram* pada metode *Naïve Bayes Classifier* mampu meningkatkan akurasi model dari yang semula 82,14% menjadi 83,14%.

Salah satu tantangan utama dalam klasifikasi terletak pada permasalahan kelas tidak seimbang (*imbalanced class*). Masalah ini dapat mengakibatkan bias dalam proses pemodelan, sehingga model cenderung mengklasifikasikan data ke dalam kelas mayoritas dan mengabaikan kelas minoritas. Metode yang sering digunakan salah satunya adalah *Synthetic Minority Oversampling Technique* (SMOTE). Penelitian yang dilakukan oleh Fajriyah *et al.* (2025) menunjukkan bahwa penerapan SMOTE pada model SVM menghasilkan akurasi yang lebih tinggi dalam mengklasifikasikan sentimen publik dibandingkan dengan model tanpa SMOTE dengan akurasi, presisi, dan *recall* sebesar 93%.

Penelitian ini dilakukan untuk menganalisis sentimen opini publik pasca Pemilu 2024 terhadap presiden dan wakil presiden terpilih ke dalam kelas negatif, positif, dan netral di media sosial X. Metode klasifikasi yang digunakan adalah pendekatan *One Versus Rest* pada *Support Vector Machine*. Teknik pembobotan TF-IDF diterapkan bersamaan dengan ekstraksi fitur berbasis *N-Gram* untuk meningkatkan representasi konteks dalam data teks yang dianalisis.

Penelitian ini diharapkan dapat mengklasifikasikan sentimen dengan lebih akurat dan sesuai karakteristik data teks berbahasa Indonesia, khususnya dalam konteks wacana politik di media sosial. Model klasifikasi terbaik yang diperoleh dapat diterapkan pada data baru dengan cakupan waktu yang lebih luas guna mendukung analisis sentimen secara berkala maupun *real time*. Model terbaik juga memiliki potensi untuk diimplementasikan pada berbagai platform media sosial lainnya di luar ruang lingkup penelitian ini, guna memperluas jangkauan serta meningkatkan akurasi pemantauan opini publik. Hasil klasifikasi sentimen dapat digunakan sebagai indikator tingkat kepercayaan masyarakat terhadap pemerintah sekaligus sebagai acuan bagi pengambil kebijakan dalam merumuskan strategi komunikasi yang lebih adaptif dan responsif terhadap dinamika opini publik.

1.2 Rumusan Masalah

Rumusan masalah yang dibahas dalam penelitian ini adalah:

1. Bagaimana penerapan pembobotan TF-IDF berbasis *N-Gram* menggunakan SVM dengan pendekatan OVR pada data teks sentimen X mengenai pasangan presiden dan wakil presiden terpilih?
2. Bagaimana kinerja klasifikasi *multiclass* analisis sentimen X mengenai pasangan presiden dan wakil presiden terpilih menggunakan pembobotan TF-IDF berbasis *N-Gram* pada metode SVM dengan pendekatan OVR?

1.3 Batasan Masalah

Terdapat beberapa batasan masalah yang ditetapkan dalam penelitian ini yaitu:

1. Data yang digunakan adalah data berbahasa Indonesia dengan kombinasi kata kunci “Prabowo” dan “Gibran” berjumlah 1.272.

2. Periode pengambilan data dari media sosial X mulai dari tanggal 15 Februari 2024 sampai tanggal 2 Oktober 2024. Periode tersebut digunakan karena sesuai dengan periode pasca Pemilu Indonesia 2024.
3. Penelitian ini menggunakan kombinasi *N-Gram* dari unigram ($N = 1$), bigram ($N = 2$), hingga trigram ($N = 3$). Kombinasi tersebut dapat dituliskan sebagai *N-Gram* (1, 3).

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Menerapkan pembobotan TF-IDF berbasis *N-Gram* menggunakan SVM dengan pendekatan OVR pada data teks sentimen X mengenai pasangan presiden dan wakil presiden terpilih.
2. Mengevaluasi kinerja klasifikasi *multiclass* analisis sentimen X mengenai pasangan presiden dan wakil presiden terpilih menggunakan pembobotan TF-IDF berbasis *N-Gram* pada metode SVM dengan pendekatan OVR.