BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah membawa perubahan signifikan terhadap cara masyarakat berinteraksi dan menyampaikan pendapat di ruang digital. Salah satu wujud perkembangan tersebut adalah media sosial, yang telah membuka ruang baru bagi masyarakat untuk mengemukakan opini dan tanggapan terhadap kebijakan publik maupun kinerja pemerintahan, termasuk melalui platform X. Data tekstual dari platform tersebut menunjukkan tren besar dalam diskusi publik dan persepsi masyarakat yang secara langsung dan berkelanjutan. Hal ini menjadikan media sosial sebagai sumber data yang relevan untuk memahami sentimen publik secara lebih luas dan mendalam.

Berkembangnya opini publik yang masif di dunia maya menimbulkan kebutuhan akan pendekatan berbasis teknologi untuk menganalisis pandangan masyarakat secara objektif. Dalam ranah sistem informasi modern, pendekatan tersebut diwujudkan melalui penerapan Natural Language Processing (NLP) yang memungkinkan komputer memahami dan mengolah bahasa manusia dalam bentuk teks. Salah satu cabang penting dari NLP adalah analisis sentimen, yaitu proses untuk mengidentifikasi opini posistif, negatif, atau netral dari data teks. Penerapannya menjadi sangat relevan dalam memahami persepsi masyarakat terhadap kebijakan pemerintah khususnya pada periode 100 hari pertama pemerintahan Prabowo Subianto yang menjadi topik hangat di media sosial. Tanpa analisis berbasis data, opini publik yang beragam dan dinamis sulit diinterpretasikan secara akurat. Oleh karena itu, penerapan analisis sentimen berbasis machine learning diperlukan untuk mengubah data teks menjadi informasi yang terklasifikasi dengan baik [1]. Pendekatan ini, memungkinkan identifikasi arah kecenderungan sentimen publik terhadap pemerintahan.

Sejumlah penelitian terdahulu telah mengkaji penerapan algoritma klasifikasi dalam analisis sentimen berbahasa Indonesia. Majid dan Nuari (2025)

membandingkan model BERT dengan algoritma klasik seperti Naïve Bayes dan SVM pada beberapa dataset multibahasa dan menemukan bahwa model klasik menunjukkan kinerja lebih baik pada data Indonesia yang tidak seimbang [2]. Sementara itu, Umar dan Nur (2022) meneliti variasi algoritma Naïve Bayes dan menegaskan bahwa pentingnya pemilihan parameter serta jenis model dalam meningkatkan akurasi analisis sentimen pada data Twitter berbahasa Indonesia [3]. Merujuk pada hasil- hasil tersebut, penelitian ini secara khusus berfokus pada perbandingan performa dua model klasifikasi, yaitu Naive Bayes dan Support Vector Machine (SVM), dalam menganalisis sentimen publik terkait periode 100 hari kerja pemerintahan Prabowo Subianto. Analisis dilakukan terhadap 1.369 tweet berbahasa Indonesia yang diperoleh melalui proses scraping. Data yang terkumpul kemudian melalui tahap pra-pemrosesan teks dan dilabeli secara manual untuk memastikan kualitas kategori sentimen. Data yang sudah siap, digunakan untuk melatih model klasifikasi, selanjutnya dievaluasi menggunakan metrik accuracy, precision, recall, dan F1-score, serta divisualisasikan melalui confusion matrix. Pendekatan komparatif ini digunakan untuk menentukan model yang paling optimal dalam mengklasifikasikan sentimen pada teks berbahasa Indonesia dalam konteks sosial-politik.

Penelitian ini memiliki urgensi dalam mendukung pengembangan sistem informasi berbasis analisis teks di Indonesia, khususnya dalam konteks evaluasi kebijakan publik melalui media sosial. Selain itu, hasil penelitian diharapkan dapat menggambarkan persepsi masyarakat terhadap kinerja pemerintahan dalam 100 hari pertama secara lebih objektif dan berbasis data. Temuan yang dihasilkan diharapkan memberikan kontribusi ilmiah sekaligus praktis bagi pengambil kebijakan, serta penerapan teknologi analisis data sosial di lingkungan pemerintahan dan akademik.

1.2 Rumusan Masalah

1. Bagaimana hasil analisis sentimen publik terhadap periode 100 hari kerja pemerintahan Prabowo Subianto berdasarkan data yang diperoleh dari platform X?

2. Di antara model klasifikasi Naive Bayes dan Support Vector Machine, model manakah yang menunjukkan performa lebih baik dalam analisis sentimen publik berdasarkan metrik evaluasi accuracy, precision, recall, dan F1-score?

1.3 Batasan Masalah

1. Data yang digunakan berasal dari platform X (Twitter) dengan jumlah 1.369 data teks berbahasa Indonesia yang diperoleh melalui proses scraping menggunakan Twitter Harvester dan Grok AI, yang terbatas pada topik 100 hari kerja pemerintahan Prabowo Subianto dan dikumpulkan pada periode Januari hingga Februari 2025. Pengambilan data dilakukan menggunakan kata kunci pencarian "100 hari kerja prabowo", "100 hari prabowo" dan "100 hari pemerintahan". Analisis hanya difokuskan pada kolom teks utama (full_text), tanpa mempertimbangkan metadata lain seperti waktu unggahan, jumlah suka, atau nama pengguna.
2. Analisis sentimen hanya mencakup tiga kategori sentimen, yaitu positif, negatif, dan netral, berdasarkan proses pelabelan manual.
3. Penelitian ini hanya membandingkan dua model klasifikasi, yaitu Multinomial Naive Bayes dan Support Vector Machine Linear (SVM).
4. Evaluasi performa model dibatasi pada metrik accuracy, precision, recall, dan F1-score, serta visualisasi confusion matrix sebagai alat interpretasi hasil klasifikasi.
5. Faktor eksternal seperti waktu pengambilan data di luar periode penelitian, perubahan opini publik, serta konteks politik yang berkembang setelah Februari 2025 tidak dianalisis lebih lanjut.

1.4 Tujuan Penelitian

1. Menganalisis sentimen publik terhadap periode 100 hari kerja pemerintahan Prabowo Subianto berdasarkan data teks dari platform X (Twitter) berbahasa Indonesia, untuk mengetahui kecenderungan

dominasi sentimen positif, negatif, atau netral.

2. Membandingkan performa dua model klasifikasi teks, yaitu Naive Bayes dan Support Vector Machine (SVM), dalam menganalisis sentimen publik menggunakan pendekatan machine learning dengan pembobotan TF-IDF dan penyeimbangan data menggunakan SMOTE.
3. Mengevaluasi hasil klasifikasi berdasarkan metrik accuracy, precision, recall, dan F1-score, serta memvisualisasikan hasil menggunakan confusion matrix.

1.5 Manfaat Penelitian

Penelitian ini memiliki manfaat yang dapat ditinjau dari dua sisi, yaitu secara teoretis dan praktis. Dari sisi teoretis, hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang analisis sentimen berbahasa Indonesia, khususnya dalam penerapan algoritma machine learning seperti Naive Bayes dan Support Vector Machine (SVM). Melalui proses perbandingan performa kedua model tersebut, penelitian ini dapat memperkaya referensi empiris mengenai efektivitas metode klasifikasi teks dengan pembobotan Term Frequency – Inverse Document Frequency (TF-IDF) serta penyeimbangan data menggunakan Synthetic Minority Oversampling Technique (SMOTE). Selain itu, penelitian ini juga dapat menjadi dasar teoretis bagi penelitian selanjutnya yang ingin mengembangkan pendekatan baru, seperti deep learning atau hybrid model, dalam konteks pemrosesan bahasa alami.

Dari sisi praktis, penelitian ini diharapkan mampu memberikan informasi berbasis data mengenai kecenderungan sentimen publik terhadap kinerja pemerintahan pada periode awal kepemimpinan. Hasil analisis ini dapat dimanfaatkan oleh pemerintah, lembaga survei, serta analis kebijakan untuk memahami persepsi masyarakat di media sosial secara lebih objektif. Selain itu, hasil penelitian dapat berfungsi sebagai bahan evaluasi bagi instansi atau tim komunikasi publik dalam menyusun strategi komunikasi yang lebih adaptif terhadap opini masyarakat.

Penelitian ini juga memberikan manfaat praktis bagi pengembang sistem informasi dan data scientist, karena dapat dijadikan sebagai acuan dalam membangun model analisis sentimen teks berbahasa Indonesia. Penggunaan kombinasi metode praproses teks, pembobotan TF-IDF, serta algoritma klasifikasi yang dibandingkan diharapkan dapat menjadi referensi teknis yang implementatif untuk berbagai kebutuhan analisis opini publik di platform media sosial.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, Berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian serta sistematika penulisan, Bab ini memberikan gambaran umum mengenai konteks penelitian dan alasan utama dilakukannya analisis komparatif algoritma.

BAB II TINJAUAN PUSTAKA, Bab ini memuat kajian teori yang relevan dengan penelitian dan mencakup tinjauan terhadap penelitian-penelitian terdahulu yang sejenis untuk memperkuat landasan teori.

BAB III METODE PENELITIAN, Bab ini menjelaskan secara rinci tahapan dan metode yang digunakan dalam penelitian, mulai dari pengumpulan data, preprocessing teks, pelabelan manual data, pembagian dataset, penanganan data tidak seimbang, penerapan algoritma klasifikasi, optimasi model dan evaluasi model.

BAB IV HASIL DAN PEMBAHASAN, Bab ini menyajikan hasil penerapan metode penelitian meliputi hasil pengumpulan data, pra-pemrosesan teks, distribusi sentimen, performa masing-masing model klasifikasi, serta perbandingan hasil evaluasi menggunakan metrik yang telah ditentukan. Bab ini juga berisi analisis terhadap hasil yang diperoleh untuk mengidentifikasi model dengan kinerja yang paling optimal dan menafsirkan kecenderungan sentimen publik berdasarkan data yang diperoleh.

BAB V PENUTUP, Bab ini berisi kesimpulan yang diperoleh dari hasil penelitian serta saran yang dapat diberikan untuk penelitian selanjutnya .