

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia menganut sistem demokrasi, di mana Pemilu menjadi sarana rakyat memilih wakil dan pejabat publik. Pemilu dilaksanakan secara berkala untuk memastikan pemerintahan sesuai kehendak rakyat. Pada 2024, Indonesia akan menggelar pemilu serentak untuk memilih Presiden dan Wakil Presiden. Sejumlah tokoh politik mulai muncul sebagai calon berdasarkan opini masyarakat, yang mencerminkan pandangan dan harapan publik terhadap calon pemimpin. Selain itu, Pilkada juga sering menjadi topik yang menimbulkan pro dan kontra di tengah masyarakat, bahkan kerap menjadi bahan diskusi hangat di media sosial seperti Twitter.[1]

Twitter/X kini menjadi platform komunikasi yang sangat populer di kalangan pengguna internet dan memungkinkan masyarakat Indonesia untuk menyuarakan pendapat mereka secara terbuka. Selain digunakan oleh masyarakat umum, Twitter juga dimanfaatkan oleh calon Presiden atau kandidat politik sebagai media kampanye untuk membangun citra diri, menyampaikan visi dan misi, serta menjangkau calon pemilih dan pendukung. Pada pemilihan umum di kawasan Asia tahun 2019, para kandidat secara aktif menggunakan media sosial khusus nya Twitter untuk menyebarkan slogan, mempromosikan kebijakan, menyerang popularitas lawan, serta menggalang dukungan menjelang masa kampanye.[2]

Namun demikian, tingginya partisipasi ini juga menimbulkan berbagai permasalahan, seperti meningkatnya polarisasi antarpengukung calon, ujaran kebencian, serta dominasi buzzer dan akun bot yang menyebarkan narasi tertentu secara masif. Permasalahan ini menjadi semakin kompleks ketika dikaitkan dengan rendahnya literasi digital sebagian pengguna, yang menyebabkan mereka mudah terpapar disinformasi tanpa kemampuan memverifikasi kebenarannya secara kritis. Oleh karena itu, dibutuhkan sebuah metode yang efisien dan akurat, salah satunya

adalah analisis sentimen, untuk mengklasifikasikan opini publik secara otomatis dan memahami arah perbincangan politik di Twitter selama masa pemilu berlangsung.[3]

Analisis sentimen merupakan salah satu jenis penelitian dalam bidang data mining yang bertujuan untuk menganalisis, mengekstrak, dan mengolah data dalam bentuk teks secara otomatis guna mengetahui opini, emosi, atau sikap seseorang terhadap suatu entitas, topik, atau peristiwa tertentu. Proses ini penting untuk mengungkap informasi berharga dari data tidak terstruktur seperti komentar, ulasan, atau cuitan di media sosial. Dalam analisis sentimen, salah satu metode yang sering digunakan untuk mengklasifikasikan opini ke dalam kategori positif, negatif, atau netral adalah algoritma *Naïve Bayes*, karena kemampuannya yang sederhana namun cukup efektif dalam menangani data teks berskala besar.[4]

Text mining merupakan salah satu bentuk dari data mining yang bertujuan untuk menemukan pola atau informasi penting dari kumpulan data berbentuk teks dalam jumlah besar. Perbedaan utama antara keduanya terletak pada jenis data yang diolah. *Text mining* bekerja dengan data teks yang tidak terstruktur, seperti artikel, komentar, atau tweet, sedangkan data mining biasanya menggunakan data yang sudah tersusun rapi dalam bentuk tabel atau format terstruktur lainnya. Teknik *text mining* dipilih karena sebagian besar data di dunia nyata, khususnya di internet dan media sosial, berbentuk teks yang tidak terstruktur. Oleh karena itu, Teknik *text mining* menjadi solusi efektif untuk menggali wawasan atau opini dari data tersebut yang tidak dapat dianalisis dengan pendekatan data mining konvensional.[3]

Naïve Bayes adalah metode klasifikasi yang didasarkan pada perhitungan peluang dan menganggap setiap fitur atau atribut dalam data tidak saling bergantung. Metode ini memiliki beberapa jenis, seperti Multinomial *Naïve Bayes* yang cocok untuk menghitung seberapa sering kata muncul dalam teks, Bernoulli *Naïve Bayes* untuk data dengan nilai ya atau tidak, dan Gaussian *Naïve Bayes* untuk data angka yang menyebar secara normal. Penggunaan *Naïve Bayes* dalam analisis sentimen sudah banyak dilakukan, misalnya untuk melihat opini masyarakat

tentang e-KTP, sistem plat nomor ganjil-genap, ulasan film, hingga tanggapan pengguna terhadap layanan seperti Lazada di media sosial.[6]

Dalam penelitian ini metode Multinomial *Naïve Bayes* dipilih karena dianggap paling sesuai untuk menganalisis opini masyarakat dan sering digunakan dalam pengolahan teks singkat yang tersebar di Twitter selama Pemilu Presiden 2024.[7] Algoritma ini menganggap bahwa kemunculan kata dalam suatu dokumen bersifat independen, tanpa memperhatikan urutan atau konteks kata tersebut. Berdasarkan penelitian sebelumnya yang berjudul "*A Comparative Analysis of Machine Learning Classifiers for Twitter Sentiment Analysis*", yang menggunakan beberapa varian algoritma seperti *Multinomial Naïve Bayes*, *Bernoulli Naïve Bayes*, dan SVM. Dibandingkan dengan *Bernoulli Naïve Bayes* dan SVM. Hasil penelitian menunjukkan bahwa *Multinomial NB* memiliki tingkat akurasi yang lebih tinggi dibandingkan dua metode lainnya.[8]

Penelitian lanjutan yang menggunakan dua jenis dataset yang pertama berisi label berdasarkan emosi, dan yang kedua berdasarkan sentimen positif dan negatif juga menunjukkan hasil serupa. Multinomial NB memperoleh akurasi sebesar 91,6% pada dataset pertama, dan 87,6% pada dataset kedua.[8] Berdasarkan hasil tersebut, algoritma Multinomial NB menunjukkan kinerja yang konsisten dan unggul dalam klasifikasi teks, sehingga dalam penelitian ini metode tersebut dipilih karena dianggap paling sesuai untuk menganalisis opini masyarakat yang tersebar dalam bentuk teks singkat di Twitter selama Pemilu Presiden 2024.

Penelitian ini memfokuskan klasifikasi hanya pada komentar yang bersentimen positif dan negatif, dengan alasan untuk menyederhanakan proses analisis serta mengarahkan fokus pada sejauh mana masyarakat memberikan dukungan atau penolakan terhadap kandidat presiden. Sentimen netral sengaja tidak dianalisis karena cenderung mengandung ambiguitas atau informasi yang kurang eksplisit dalam menunjukkan sikap terhadap topik politik tertentu, sehingga dapat mengurangi akurasi dan interpretabilitas hasil analisis. Dengan membatasi sentimen pada dua kategori utama, hasil klasifikasi dapat lebih jelas digunakan

sebagai dasar pemetaan opini publik yang tegas terhadap calon presiden selama masa kampanye di media sosial.

Sebagai bentuk validasi metode, dilakukan pengujian model klasifikasi sentimen menggunakan algoritma Multinomial *Naïve Bayes*. Tweet bertema politik terkait Pemilu Presiden 2024 dikumpulkan berdasarkan kata kunci tertentu, lalu diproses melalui tahap pembersihan dan normalisasi teks seperti *case folding*, *tokenisasi*, *stopword removal*, dan *stemming*. Setelah diberi label sentimen positif atau negatif, data diubah menjadi vektor menggunakan metode *Bag-of-Words* atau *TF-IDF*. Dataset dibagi menjadi data latih (80%) dan data uji (20%), kemudian digunakan untuk melatih dan menguji model. Evaluasi kinerja dilakukan menggunakan metrik akurasi, presisi, *recall*, dan *F1-score* untuk mengukur efektivitas model dalam mengklasifikasikan opini publik di Twitter selama masa kampanye.[9]

Penelitian ini bertujuan mengembangkan model klasifikasi sentimen dengan algoritma Multinomial *Naïve Bayes* untuk mengidentifikasi opini publik di Twitter terkait Pemilu Presiden 2024 dalam bentuk positif dan negatif. Hasil yang diharapkan adalah terciptanya model yang akurat dalam menganalisis sentimen politik serta memberikan gambaran arah dukungan atau penolakan masyarakat terhadap calon presiden. Berbeda dengan penelitian sebelumnya yang menggunakan optimasi seperti PSO, seleksi fitur *Chi-Square*, atau algoritma SVM dengan *Word2Vec*, penelitian ini fokus pada Multinomial *Naïve Bayes* dengan pembobotan *TF-IDF* dan penyeimbangan data menggunakan SMOTE. Pendekatan ini dinilai lebih sesuai dengan karakteristik data Twitter yang singkat dan tidak seimbang, sehingga diharapkan menghasilkan model yang lebih stabil dan konsisten. Penelitian ini diharapkan bermanfaat bagi pemerintah, media, dan masyarakat dalam memahami persepsi publik secara data-driven serta meningkatkan kesadaran literasi digital di tengah arus informasi politik di media sosial.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan rumusan masalah penelitian ini sebagai berikut:

1. Bagaimana proses penerapan algoritma *Multinomial Naïve Bayes* dalam melakukan analisis sentimen terhadap data Twitter terkait Pemilu 2024?
2. Bagaimana hasil evaluasi model berdasarkan *confusion matrix*, meliputi nilai akurasi, presisi, *recall*, dan *F1-score*?

1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Data yang digunakan dalam penelitian ini diambil dari aplikasi X (Twitter) dengan kata kunci yang berkaitan dengan hasil pemilu 2024
2. Analisis sentiment hanya dikategorikan ke dalam 2 kelas yaitu positif dan negative
3. Metode klasifikasi yang digunakan terbatas pada algoritma *Multinomial Naïve bayes*, tanpa dibandingkan dengan algoritma lain.
4. Data yang dianalisis hanya berupa tweet dalam Bahasa Indonesia, dan tidak mempertimbangkan unsur sarkasme, ironi, atau konteks budaya secara mendalam.
5. Preprocessing data yang dilakukan mencakup proses pembersihan teks seperti penghapusan tanda baca, *stopword*, *URL*, dan normalisasi kata.

1.4 Tujuan Penelitian

Penelitian ini bertujuan mengembangkan model analisis sentimen dengan algoritma *Multinomial Naïve Bayes* untuk mengklasifikasikan opini masyarakat di Twitter terkait Pemilu Presiden 2024. Model ini diharapkan mampu memberikan gambaran arah dukungan atau penolakan publik secara akurat dan relevan sebagai rujukan dalam memahami dinamika politik di ruang.

Menerapkan dan mengevaluasi algoritma *Multinomial Naive Bayes* dalam

klasifikasi sentimen terhadap data Twitter mengenai Pemilu Presiden 2024, dengan menilai performa model menggunakan metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score* berdasarkan *confusion matrix*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pemahaman opini publik terkait Pemilu Presiden 2024 melalui pendekatan analisis sentimen berbasis algoritma Multinomial *Naive Bayes*. Dari sisi teknis, penelitian ini bermanfaat dalam menunjukkan proses penerapan dan evaluasi model klasifikasi sentimen, khususnya dengan menyajikan hasil pengukuran akurasi, presisi, *recall*, dan *F1-score* melalui *confusion matrix*, sehingga dapat menjadi referensi dalam pengembangan sistem analisis opini publik secara otomatis.

Bagi pihak seperti pemerintah, penyelenggara pemilu, dan media, hasil klasifikasi ini dapat digunakan untuk memantau persebaran sentimen positif dan negatif terhadap calon presiden secara berbasis data dan *real-time*. Informasi tersebut juga dapat membantu dalam mengidentifikasi tren komunikasi, mendeteksi potensi disinformasi, serta merancang strategi komunikasi politik yang lebih efektif dan responsif.

Selain itu, bagi masyarakat umum, penelitian ini diharapkan dapat menjadi sarana edukatif yang mendorong peningkatan literasi digital, khususnya dalam mengenali opini yang valid dan membedakannya dari narasi yang manipulatif di media sosial selama masa pemilu.

1.6 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian tugas akhir ini dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan berbagai komponen penting dalam penelitian, yang meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan sebagai panduan dalam penyusunan isi skripsi secara terstruktur.

BAB II TINJAUAN PUSTAKA

Bab ini memuat tinjauan pustaka dari penelitian-penelitian terdahulu yang dijadikan referensi oleh penulis. Selain itu, bab ini juga membahas landasan teori dan konsep-konsep yang berkaitan langsung dengan permasalahan yang diteliti dalam penelitian ini.

BAB III METODE PENELITIAN

Bab ini membahas mengenai metode yang digunakan dalam penelitian, yang mencakup teknik pengumpulan data serta metode yang digunakan untuk menganalisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas hasil penelitian analisis sentimen terkait Pemilu Presiden 2024 yang telah dilakukan menggunakan metode yang telah ditetapkan sebelumnya.

BAB V PENUTUP

Bab ini berisi kesimpulan yang diperoleh berdasarkan rumusan masalah yang telah disusun, serta berisi saran untuk penelitian selanjutnya.