

BAB I

PENDAHULUAN

1.1 Latar Belakang

Proyek pembangunan Ibu Kota Nusantara (IKN) adalah agenda nasional yang otomatis menarik perhatian luas dan memicu berbagai respons di masyarakat. Opini dan tanggapan dari warganet, terutama yang beredar di media sosial, menjadi cerminan penting mengenai penerimaan dan potensi tantangan proyek ini ke depan. Platform Twitter (X), karena sifatnya yang cepat dan jangkauannya yang besar, menjadi tempat utama bagi munculnya beragam sentimen, baik yang sifatnya mendukung maupun kritis [1], [2]. Kumpulan data opini yang sangat besar ini memerlukan sebuah metode yang sistematis untuk bisa mengidentifikasi arah dan kecenderungan sentimen secara akurat dan efisien, karena analisis manual menjadi tidak efektif pada volume data yang tinggi [3].

Fokus penelitian ini adalah menganalisis sentimen warganet yang terkait spesifik dengan isu IKN pada sektor pendidikan, yang mana sektor ini sangat penting bagi masa depan kualitas sumber daya manusia dan infrastruktur ilmu pengetahuan. Data tekstual yang diamati menunjukkan adanya perbedaan sudut pandang (polarisasi) dan variasi (heterogenitas) yang tinggi dalam sentimen yang disuarakan. Ada kesulitan signifikan untuk menarik kesimpulan pandangan umum secara cepat dan objektif apabila hanya mengandalkan proses analisis manual [1]. Oleh karena itu, kondisi data yang tidak stabil ini menuntut adanya solusi berbasis komputasi untuk gambaran dan mengukur kestabilan sentimen.

Guna mengatasi kesulitan analisis data teks yang volumenya besar, digunakanlah teknik klasifikasi sentimen menggunakan machine learning. Algoritma seperti Naive Bayes (NB) dan Support Vector Machine (SVM) dikenal memiliki potensi perbandingan yang baik dalam tugas klasifikasi teks [4] dan telah banyak diterapkan serta diperbandingkan dalam analisis

sentimen isu IKN di media sosial [5], [2]. Peneliti melihat adanya kebutuhan mendesak untuk membandingkan secara objektif dan terukur performa dari kedua algoritma tersebut. Perbandingan ini bertujuan untuk menemukan konfigurasi kinerja terbaik yang sanggup menghasilkan tingkat akurasi tertinggi saat memilih dan mengelompokkan sentimen warganet.

Luaran dari kegiatan penelitian ini diharapkan dapat menyajikan gambaran yang terukur dan valid mengenai situasi sentimen publik IKN khususnya di sektor pendidikan. Temuan yang dihasilkan dapat menjadi data masukan penting bagi pihak-pihak terkait, khususnya dalam perumusan strategi komunikasi serta pengambilan keputusan yang berlandaskan bukti pengalaman dan pengamatan. Dengan menyediakan alat analisis yang teruji dan dapat dipercaya, penelitian ini menegaskan relevansinya dan menjadi kontribusi esensial dalam mendukung pemahaman data publik sesuai dengan fokus judul yang diangkat.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, masalah-masalah yang akan diselesaikan dalam penelitian skripsi ini dirumuskan dalam bentuk kalimat tanya sebagai berikut:

1. Bagaimana perbandingan hasil klasifikasi sentimen warganet Twitter mengenai Ibu Kota Nusantara (IKN) di sektor pendidikan menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM)?
2. Berapakah tingkat akurasi tertinggi yang dihasilkan oleh algoritma klasifikasi sentimen optimal (Naive Bayes atau SVM) terkait isu IKN pada sektor pendidikan?

1.3 Batasan Masalah

Agar penelitian ini dapat berjalan fokus dan menghasilkan temuan yang relevan dengan tujuan, berikut adalah batasan-batasan masalah yang ditetapkan:

1. Data yang digunakan berupa tweet yang diperoleh dari platform Twitter (X) melalui proses *crawling* menggunakan kata kunci yang berkaitan dengan Ibu Kota Nusantara (IKN) dan sektor pendidikan. Pengambilan data dilakukan pada periode Oktober – Desember 2025 dengan total data yang dianalisis sebanyak 2006 tweet, sehingga hasil penelitian ini merepresentasikan opini publik pada rentang waktu tersebut.
2. Analisis hanya difokuskan pada sentimen masyarakat terhadap pembangunan IKN yang berkaitan dengan sektor pendidikan, seperti sekolah, guru, kampus, fasilitas, dan kebijakan pendidikan. Isu di luar sektor pendidikan tidak termasuk dalam pembahasan.
3. Data yang dianalisis terbatas pada tweet berbahasa Indonesia dan berbentuk teks, tanpa mempertimbangkan gambar, video, atau bentuk konten lainnya.
4. Proses pelabelan sentimen dilakukan secara otomatis menggunakan model IndoBERT sebagai alat bantu klasifikasi awal. Penelitian ini tidak melakukan validasi manual terhadap seluruh data karena keterbatasan waktu dan sumber daya.
5. Penelitian ini hanya membandingkan dua algoritma klasifikasi, yaitu Multinomial Naive Bayes dan Support Vector Machine (SVM), dengan representasi fitur menggunakan metode TF-IDF. Pengujian tidak mencakup algoritma lain maupun pendekatan *deep learning*.
6. Penelitian ini tidak menganalisis faktor eksternal seperti keberadaan akun bot, tingkat keaktifan pengguna, lokasi geografis, maupun karakteristik demografis pengguna Twitter.

7. Hasil penelitian ini tidak dimaksudkan untuk menggeneralisasi seluruh opini masyarakat Indonesia, melainkan hanya menggambarkan pola sentimen berdasarkan data yang diperoleh pada periode penelitian.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian berjudul Komparasi Kinerja Algoritma Naive Bayes dan Support Vector Machine dalam Klasifikasi Sentimen Warganet Twitter Mengenai Isu Ibu Kota Nusantara pada Sektor Pendidikan adalah:

1. Menganalisis perbandingan kinerja antara algoritma Naive Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen warganet Twitter mengenai isu Ibu Kota Nusantara pada sektor pendidikan.
2. Mengetahui tingkat akurasi terbaik yang dihasilkan dalam klasifikasi sentimen warganet mengenai isu IKN pada sektor pendidikan, serta menentukan algoritma mana yang lebih optimal antara Naive Bayes dan SVM.

1.5 Manfaat Penelitian

Kegiatan penelitian ini diharapkan dapat memberikan manfaat yang signifikan, terukur, dan berkelanjutan, baik dari segi pengembangan ilmu pengetahuan (manfaat teoritis) maupun kontribusi nyata bagi pihak-pihak terkait (manfaat praktis), sehingga penelitian ini dapat berkontribusi pada pengembangan ilmu teknologi informasi serta memberikan dampak positif bagi pemangku Kepentingan IKN Sektor Pendidikan dan pengguna teknologi. Adapun manfaat yang diperoleh dari penelitian ini, adalah sebagai berikut:

1. Manfaat Teoritis
 - a. Pengembangan Ilmu Komputasi dan NLP: Penelitian ini menyumbang data nyata dari pengalaman yang mendalam mengenai

perbandingan komposisi kinerja algoritma Naive Bayes dan Support Vector Machine (SVM) dalam konteks klasifikasi sentimen pada data teks Bahasa Indonesia yang terpolarisasi. Hasilnya akan memperkaya literatur dan menjadi referensi valid bagi studi machine learning dan Natural Language Processing (NLP) di Indonesia, terutama yang berfokus pada isu sosial-politik sensitif.

- b. Validasi Model untuk Data Tidak Stabil: Hasil penelitian ini dapat memvalidasi dan mengidentifikasi algoritma mana yang lebih tangguh dan stabil dalam menghadapi data sentimen yang memiliki tingkat polarisasi dan heterogenitas tinggi, seperti yang terjadi pada isu IKN. Ini memberikan panduan metodologis bagi peneliti selanjutnya saat memilih model klasifikasi untuk data sosial yang dinamis.
- c. Dasar Penelitian Lanjutan: Temuan terkait tingkat akurasi, presisi, dan recall serta konfigurasi optimal dari algoritma terpilih (Naive Bayes atau SVM) dapat digunakan sebagai landasan dan tolok ukur yang kuat bagi peneliti berikutnya yang ingin mengembangkan model sentimen yang lebih canggih.

2. Manfaat Praktis

- a. Bagi Pemangku Kepentingan IKN Sektor Pendidikan: Menyediakan gambaran yang terukur, cepat, dan objektif mengenai penerimaan publik terhadap isu IKN pada sektor pendidikan. Data hasil klasifikasi sentimen ini berfungsi sebagai early warning system yang krusial, membantu pihak IKN, Kementerian Pendidikan, atau instansi terkait lainnya dalam perumusan kebijakan dan strategi komunikasi
- b. Bagi Pengguna Teknologi Klasifikasi Sentimen: Menghasilkan model klasifikasi yang optimal dan teruji yang dapat diimplementasikan dalam pemantauan sentimen real-time. Model ini

memberikan solusi komputasi yang efisien untuk mengatasi volume data yang besar dan tidak stabil, sehingga proses pemantauan opini publik menjadi lebih cepat dan akurat.

- c. Bagi Komunitas Akademik dan Praktisi Data: Menyediakan dataset sentimen IKN Sektor Pendidikan (yang telah dilabeli) sebagai sumber daya terbuka (jika diizinkan) yang dapat dimanfaatkan oleh mahasiswa, peneliti, atau praktisi data untuk tujuan edukasi, pengujian model baru, atau replikasi penelitian.

1.6 Sistematika Penulisan

1. BAB I PENDAHULUAN

Berisi Latar Belakang Masalah, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, dan Manfaat Penelitian. Bab ini menjelaskan alasan dilakukannya penelitian, fokus isu sentimen (IKN Sektor Pendidikan), perbandingan algoritma Naive Bayes dan Support Vector Machine (SVM), serta ruang lingkup penelitian yang dibatasi.

2. BAB II TINJAUAN PUSTAKA

Berisi Studi Literatur dari penelitian terdahulu yang relevan dan Dasar Teori yang digunakan. Dasar-dasar teori ini memuat ilmu-ilmu dasar yang relevan dengan parameter penelitian dan berfungsi sebagai landasan untuk menyelesaikan masalah penelitian.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan secara rinci proses dan tahapan yang dilakukan dalam penelitian. Alur Penelitian (prosedur langkah-langkah yang diilustrasikan melalui *flowchart*), Alat dan Bahan yang mencakup data penelitian, alat/instrumen yang digunakan, dan solusi yang ditawarkan.

4. BAB IV HASIL DAN PEMBAHASAN

BAB IV berisi hasil pengolahan data, proses pelabelan, ekstraksi fitur TF-IDF, pemodelan menggunakan Naive Bayes dan Support Vector Machine, serta evaluasi dan perbandingan kinerja kedua algoritma.

5. BAB V PENUTUP

Berisi Kesimpulan yang merupakan intisari dari pembahasan dan harus menjawab rumusan masalah, serta Saran yang berisi hal-hal yang masih dapat dikembangkan lebih lanjut atau berisi masalah-masalah yang dialami pada saat proses pengerjaan skripsi.

