

Article history:
Received: Jul 9, 2025
Published: Dec 6, 2025

PERSEPSI PUBLIK TERHADAP PERPINDAHAN IBU KOTA KE IKN BERDASARKAN ANALISIS SENTIMEN X MENGGUNAKAN NAÏVE BAYES CLASSIFIER

Fauzan¹⁾, Wiwik Kusrini²⁾, Fathurrahmani³⁾

^{1, 2, 3} Program Studi Teknologi Informasi, Jurusan Komputer dan Bisnis, Politeknik Negeri Tanah Laut
email: fauzanes22@gmail.com, wiwik.kusrini@politala.ac.id, fathurrahmani@politala.ac.id

Abstract

The relocation of Indonesia's capital city from Jakarta to Nusantara (IKN) has become a topic that generates various public reactions. These opinions are largely expressed through the social media platform X (formerly Twitter), making it an important data source for analyzing public sentiment. This study aims to apply the Naïve Bayes Classifier method to analyze public sentiment toward the relocation of IKN and to evaluate its accuracy using a Confusion Matrix. The data were collected through a web crawling process on the X platform using the keyword "ikn", resulting in 1,002 relevant tweets. The data were then processed through preprocessing stages using Python, including lowercasing, punctuation removal, tokenizing, stopword removal, stemming, and TF-IDF value calculation to form features. After augmentation, the number of data increased to 1,324, which was divided into 75% training data and 25% testing data. Sentiment classification was performed using the Naïve Bayes Classifier algorithm, and testing was conducted using the black-box method to ensure system functionality according to the design. Model performance evaluation using a Confusion Matrix produced an accuracy of 92.47%, precision of 93.03%, recall of 92.47%, and F1-score of 92.45%. The web-based system developed in this study is expected to help the government understand public opinion and develop more effective communication strategies related to the IKN relocation policy.

Keywords: Confusion Matrix, IKN, Naïve Bayes Classifier, Sentiment Analysis, X

Abstrak

Perpindahan Ibu Kota Negara dari Jakarta ke Nusantara (IKN) menjadi topik yang menimbulkan beragam reaksi di kalangan masyarakat. Berbagai opini publik disampaikan melalui platform media sosial X (sebelumnya Twitter), sehingga menjadi sumber data penting untuk menganalisis sentimen masyarakat. Penelitian ini bertujuan untuk menerapkan metode *Naïve Bayes Classifier* dalam menganalisis sentimen publik terhadap perpindahan IKN serta mengevaluasi tingkat akurasi menggunakan *Confusion Matrix*. Data dikumpulkan melalui proses *web crawling* pada platform X dengan kata kunci "ikn", yang menghasilkan 1.002 *tweet* yang relevan. Data kemudian diproses melalui tahapan *preprocessing* menggunakan Python, meliputi *lowercase*, penghapusan tanda baca, *tokenizing*, *stopword removal*, dan *stemming*, serta perhitungan nilai *TF-IDF* untuk membentuk fitur. Setelah tahap *augmentasi*, jumlah data bertambah menjadi 1.324, yang dibagi menjadi 75% data latih dan 25% data uji. Klasifikasi sentimen dilakukan menggunakan algoritma *Naïve Bayes Classifier*, dan pengujian dilakukan dengan metode *black-box* untuk memastikan fungsionalitas sistem berjalan sesuai rancangan. Evaluasi performa model menggunakan *Confusion Matrix* menghasilkan akurasi sebesar 92,47%, presisi 93,03%, *recall* 92,47%, dan *f1-score* 92,45%. Sistem yang dikembangkan berbasis web ini diharapkan dapat membantu pemerintah memahami opini masyarakat dan menyusun strategi komunikasi yang lebih efektif terkait kebijakan perpindahan IKN.

Kata Kunci: Analisis Sentimen, *Confusion Matrix*, IKN, *Naïve Bayes Classifier*, X

1. PENDAHULUAN

Pemerintah Indonesia secara resmi mengumumkan rencana strategis untuk memindahkan Ibu Kota Negara dari Jakarta ke Nusantara (IKN) di Kalimantan Timur pada tanggal 26 Agustus 2019, berdasarkan Keputusan Presiden Nomor 63 Tahun 2019. Kebijakan ini diambil sebagai solusi terhadap berbagai permasalahan krusial yang dihadapi Jakarta, seperti kepadatan penduduk, kemacetan lalu lintas, banjir berkepanjangan, dan tingkat polusi yang tinggi. Dengan memindahkan pusat pemerintahan ke IKN, diharapkan terjadi pemerataan pembangunan nasional dan peningkatan kualitas hidup di wilayah luar Pulau Jawa.

Kebijakan besar ini memicu beragam tanggapan dari masyarakat yang dengan cepat menyebar melalui berbagai platform media sosial. Salah satu platform yang paling banyak digunakan untuk menyuarakan opini publik adalah Twitter, yang kini dikenal sebagai X. Media sosial X memiliki karakteristik terbuka dan cepat dalam penyebaran informasi, memungkinkan siapa pun untuk menyampaikan pendapat tanpa harus saling mengikuti akun satu sama lain, berbeda dengan media sosial seperti Facebook atau Instagram. Kemampuannya dalam menampilkan isu-isu trending menjadikan X sebagai indikator opini publik yang dinamis dan responsif. Dengan lebih dari 140 juta pengguna aktif dan ratusan juta tweet yang diposting setiap hari, platform ini menyediakan kumpulan data teks dalam jumlah besar yang potensial untuk diteliti. Selain itu, X juga menyediakan API terbuka yang memungkinkan peneliti melakukan pengumpulan data secara otomatis berdasarkan kata kunci, tagar, atau lokasi tertentu [1].

Pengumpulan opini publik dari tweet dapat dilakukan secara langsung, namun menganalisis data dalam jumlah besar secara manual tentu sangat tidak efisien. Diperlukan pendekatan yang lebih cepat dan otomatis untuk mengklasifikasikan sentimen yang terkandung dalam tweet, apakah bernada positif atau negatif. Hal ini memungkinkan pemerintah untuk memperoleh gambaran yang lebih akurat mengenai persepsi masyarakat terhadap kebijakan pemindahan IKN.

Penelitian ini menerapkan algoritma *Naïve Bayes Classifier* (NBC), sebuah metode berbasis probabilitas yang unggul dalam klasifikasi teks. NBC bekerja berdasarkan Teorema Bayes dan terbukti mampu memprediksi kelas data dengan efisien dan akurat. NBC menghasilkan akurasi tertinggi (100%) dibanding metode K-Nearest Neighbor (KNN) sebesar 98,25% dan Decision

Tree sebesar 62,28% [2]. Selain itu, NBC juga dapat memberikan hasil yang cepat dan akurat pada dataset besar, dengan performa yang sebanding dengan Decision Tree dan Neural Network [3]. Berdasarkan keunggulan tersebut, NBC dipilih dalam penelitian ini untuk mengklasifikasikan sentimen publik terhadap isu pemindahan IKN menggunakan data dari X.

Perancangan sistem analisis sentimen yang efektif untuk mengolah data dari X menjadi penting dalam konteks ini. Sistem yang dibangun bertujuan untuk mengelompokkan opini publik ke dalam dua kategori sentimen, yakni positif dan negatif. Selain itu, penelitian ini juga bertujuan untuk mengukur performa algoritma dengan menggunakan *Confusion Matrix* sebagai alat evaluasi terhadap akurasi dan ketepatan prediksi sistem yang dikembangkan.

Sejumlah penelitian terdahulu telah membuktikan pentingnya analisis sentimen dalam menggambarkan opini publik. Salah satu penelitian menggunakan NBC untuk menganalisis tren olahraga selama pandemi COVID-19 dengan data dari X dan membagi data menjadi data latih serta data uji [4]. Teknik pengumpulan data secara otomatis (*crawling*) menggunakan Python juga telah dikembangkan [5]. Selain itu, penggunaan *Confusion Matrix* dalam mengevaluasi model klasifikasi telah diterapkan dalam penelitian lain [6]. Penelitian lain juga menunjukkan bahwa NBC efektif dalam menganalisis sentimen pembelajaran daring [1]. Secara keseluruhan, studi-studi tersebut mendukung validitas pemanfaatan NBC dalam konteks media sosial dan klasifikasi sentimen, sehingga mendasari penelitian ini.

TINJAUAN PUSTAKA

Analisis Sentimen

Analisis sentimen merupakan metode dalam bidang *Natural Language Processing* (NLP) dan *Machine Learning* (ML) yang digunakan untuk secara otomatis mengenali opini atau emosi dalam sebuah teks, baik positif, negatif, maupun netral. Proses ini melibatkan identifikasi kata atau frasa yang mencerminkan sikap tertentu, dengan tujuan mengungkap pemikiran dan perasaan dari data teks terstruktur maupun tidak terstruktur [7].

Dengan meningkatnya penggunaan internet, masyarakat semakin sering menyampaikan pendapat mereka di berbagai platform media sosial, salah satunya adalah Twitter. Twitter menjadi sumber data yang kaya untuk analisis sentimen karena memungkinkan pengguna mengekspresikan opini secara terbuka melalui unggahan singkat, lengkap dengan emoji dan tagar

yang mempercepat penyebaran informasi. Ciri-ciri ini menjadikan Twitter sangat relevan dalam studi analisis sentimen karena dapat merepresentasikan opini publik secara real-time dan langsung [7].

Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF merupakan metode yang digunakan dalam pengolahan teks untuk mengukur tingkat kepentingan sebuah kata dalam suatu dokumen. Model ini menggabungkan dua konsep utama, yaitu *Term Frequency* (TF), yang menunjukkan seberapa sering suatu kata muncul dalam dokumen, dan *Inverse Document Frequency* (IDF), yang mengukur seberapa jarang kata tersebut muncul dalam seluruh koleksi dokumen. Perhitungan IDF dilakukan menggunakan rumus yang ditunjukkan pada Persamaan (1).

$$idf = \log \frac{N}{df} \quad (1)$$

Di mana N adalah jumlah total dokumen dan df adalah jumlah dokumen yang mengandung term tertentu. Selanjutnya, nilai bobot suatu kata terhadap dokumen dihitung menggunakan rumus TF-IDF yang ditunjukkan pada Persamaan (2):

$$w_{dt} = TF_{dt} \times IDF_t \quad (2)$$

Nilai TF-IDF ini digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen tersebut. Metode ini banyak diterapkan dalam pencarian informasi, klasifikasi dokumen, dan analisis sentimen [8].

Naïve Bayes Classifier (NBC)

Naïve Bayes Classifier (NBC) merupakan algoritma klasifikasi berbasis probabilitas yang sederhana dan efisien. Algoritma ini menggunakan prinsip Teorema Bayes untuk memprediksi kelas dari suatu data berdasarkan probabilitas sebelumnya (*prior*) dan bukti yang diamati (*likelihood*) dengan asumsi bahwa setiap fitur bersifat independen. NBC cocok untuk analisis data teks berskala besar karena kemampuannya dalam memberikan hasil yang akurat dengan pemrosesan ringan [9]. Perhitungan probabilitas posterior dilakukan menggunakan rumus Teorema Bayes yang ditunjukkan pada Persamaan (3).

$$P(C_i|X) = \frac{P(X|C_i) \times P(C_i)}{P(X)} \quad (3)$$

Apabila terdapat kemungkinan nilai probabilitas nol, maka digunakan teknik *Laplace Smoothing* seperti pada Persamaan (4).

$$P(x_k|C_i) = \frac{\text{jumlah data } C_i \text{ dengan } x_k + 1}{\text{jumlah data kelas } C_i + |\text{nilai unik } A_k|} \quad (4)$$

Probabilitas prior suatu kelas dihitung menggunakan Persamaan (5).

$$P(C_i) = \frac{|C_i D|}{|D|} \quad (5)$$

Kelas yang diprediksi adalah kelas dengan nilai probabilitas tertinggi, sebagaimana dinyatakan dalam Persamaan (6).

$$H = \operatorname{argmax}_{C_i} P(C_i) \prod_{k=1}^n P(x_k|C_i) \quad (6)$$

Dengan pendekatan ini, NBC menjadi metode yang handal untuk pengklasifikasian teks, termasuk dalam konteks analisis sentimen publik di media sosial [10].

Confusion Matrix

Confusion Matrix merupakan metode evaluasi performa model klasifikasi dalam bentuk tabel yang menunjukkan jumlah prediksi yang benar dan salah berdasarkan kategori aktual dan prediksi [11].

Tabel 1 Confusion Matrix

Nilai Prediksi	Nilai Aktual	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Tabel 1 terdiri dari empat komponen utama yaitu True Positive (TP) yang merupakan data aktual positif dan diprediksi positif, False Positive (FP) yaitu data aktual negatif yang salah diprediksi positif, True Negative (TN) yaitu data aktual negatif dan diprediksi negatif, serta False Negative (FN) yaitu data aktual positif yang salah diprediksi negatif [11]. Dari keempat komponen tersebut dapat dihitung beberapa metrik evaluasi yang penting dalam mengukur performa model. Akurasi mengukur proporsi prediksi yang benar terhadap total keseluruhan prediksi dan dihitung menggunakan Persamaan (7).

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (7)$$

Presisi menunjukkan tingkat ketepatan model dalam memprediksi data positif dibandingkan seluruh data yang diprediksi positif. Nilai ini dihitung dengan Persamaan (8).

$$\text{Presisi} = \frac{TP}{(TP+FP)} \times 100\% \quad (8)$$

Recall atau sensitivitas mengukur kemampuan model dalam mendeteksi data positif secara tepat dan dihitung menggunakan Persamaan (9).

$$\text{Recall} = \frac{TP}{(TP+FN)} \times 100\% \quad (9)$$

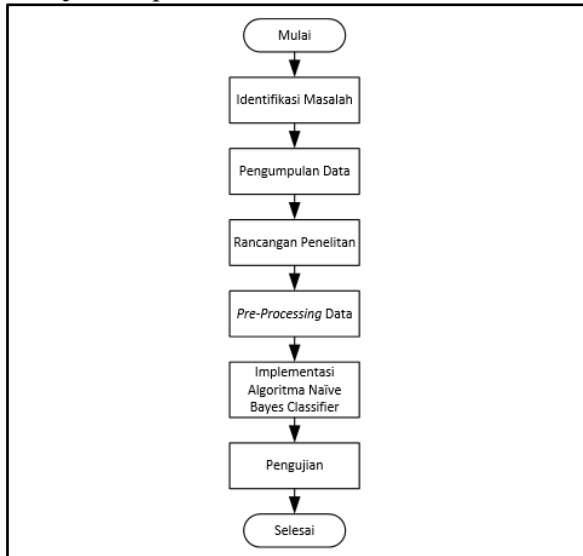
F1-Score digunakan untuk memperoleh keseimbangan antara nilai presisi dan recall dan dihitung menggunakan Persamaan (10).

$$F1 - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \times 100\% \quad (10)$$

2. METODE PENELITIAN

Rancangan Kegiatan

Penelitian ini mengikuti alur kegiatan yang tersusun secara sistematis sebagaimana ditunjukkan pada Gambar 1.



Gambar 1 Diagram Alir

Tahap pertama dimulai dari identifikasi masalah, yaitu menganalisis persepsi publik terhadap pemindahan ibu kota ke IKN. Selanjutnya dilakukan pengumpulan data dari media sosial X menggunakan teknik crawling. Setelah data terkumpul, dilakukan perancangan sistem yang melibatkan pembuatan arsitektur sistem dan antarmuka pengguna. Tahap berikutnya adalah pre-processing data, yang mencakup pembersihan dan pengolahan data mentah agar siap untuk dianalisis. Setelah itu, data dianalisis menggunakan algoritma Naive Bayes Classifier untuk mengklasifikasikan sentimen. Terakhir, dilakukan pengujian sistem dan evaluasi kinerja model menggunakan confusion matrix serta pengujian fungsionalitas dengan metode black-box testing.

Ruang Lingkup dan Objek Penelitian

Penelitian ini difokuskan pada persepsi publik terhadap pemindahan ibu kota ke Ibu Kota Negara (IKN), dengan data yang diambil dari platform media sosial X. Objek penelitian berupa unggahan teks yang mengandung kata kunci “ikn”, yang mencerminkan opini publik.

Bahan dan Alat Utama

Perangkat keras yang digunakan adalah laptop Acer Swift SF314-43 dengan spesifikasi sistem operasi Windows 11, prosesor AMD Ryzen 5, SSD 256 GB, dan RAM 16 GB. Perangkat lunak pendukung terdiri dari Visual Studio Code,

XAMPP, PHP versi 8.2.4, MySQL versi 10.4.28, web browser (Google Chrome dan Brave), Figma, Microsoft Visio 2013, dan Microsoft Office 2021.

Tempat Penelitian

Penelitian dilaksanakan di Laboratorium Bill Gates, Gedung Teknik Informatika, Politeknik Negeri Tanah Laut yang berlokasi di Jl. Ahmad Yani Km.06, Desa Panggung, Kecamatan Pelaihari, Kabupaten Tanah Laut, Kalimantan Selatan.

Teknik Pengumpulan Data

Data dikumpulkan melalui proses crawling dari media sosial X dengan kata kunci "ikn" dan menghasilkan 1002 data. Data tambahan dikumpulkan melalui studi pustaka dan wawancara guna memperkaya konteks bahasa yang digunakan dalam analisis sentimen.

Definisi Operasional Variabel Penelitian

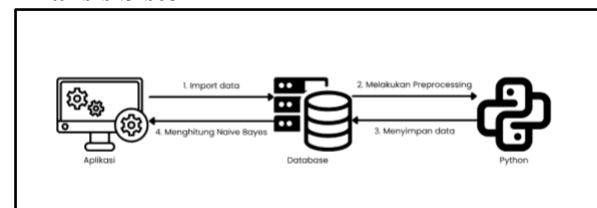
Variabel utama adalah sentimen publik yang diklasifikasikan menjadi dua kategori yaitu positif dan negatif. Data teks mentah diolah melalui tahapan pre-processing seperti lowercase, penghapusan tanda baca, tokenisasi, stopword removal, dan stemming. Kemudian bobot kata dihitung dengan metode TF-IDF untuk digunakan dalam proses klasifikasi.

Teknik Analisis

Analisis dilakukan menggunakan algoritma Naive Bayes Classifier yang mengklasifikasikan data berdasarkan probabilitas. Evaluasi kinerja model dilakukan menggunakan confusion matrix dengan mengukur akurasi, presisi, recall, dan F1-score. Selain itu, pengujian fungsionalitas sistem dilakukan dengan pendekatan black-box testing untuk memastikan sistem bekerja sesuai dengan spesifikasi.

3. HASIL DAN PEMBAHASAN

Analisis Sistem



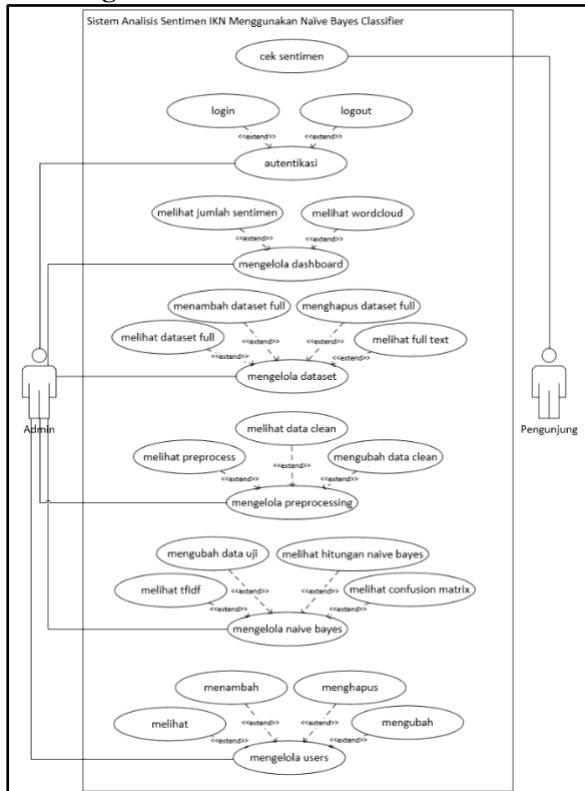
Gambar 2 Analisis Sistem

Gambar 2 menunjukkan diagram alir sistem analisis sentimen publik terhadap perpindahan ibu kota ke IKN menggunakan metode *Naive Bayes Classifier*. Sistem terdiri dari tiga komponen utama, yaitu aplikasi web, database, dan modul

Python. Pengguna mengimpor data tweet melalui antarmuka aplikasi, yang kemudian disimpan dalam database dan diproses oleh modul Python melalui tahapan *preprocessing* seperti pembersihan teks dan tokenisasi. Data hasil *preprocessing* dikembalikan ke database, lalu diklasifikasikan oleh aplikasi ke dalam dua kategori sentimen, yakni positif dan negatif. Seluruh proses berlangsung secara otomatis untuk meningkatkan efisiensi dan akurasi sistem.

(sebelumnya Twitter). Teknik pengambilan data menggunakan Python dengan bantuan Twitter API dan kata kunci “ikn” untuk menyaring topik yang relevan dengan perpindahan Ibu Kota Negara (IKN). Crawling dilakukan pada tanggal 7 Mei 2025 dan menghasilkan 1.002 tweet. Data ini digunakan sebagai dataset utama dalam proses klasifikasi sentimen menggunakan algoritma Naïve Bayes Classifier. Data terdiri dari berbagai atribut, namun hanya kolom teks (*full_text*) yang digunakan sebagai objek utama pengolahan.

Rancangan Sistem



Gambar 3 Rancangan Sistem

Gambar 3 memperlihatkan *Use Case Diagram* sistem analisis sentimen IKN dengan algoritma *Naïve Bayes Classifier*, yang melibatkan dua aktor, yaitu Admin dan Pengunjung. Pengunjung hanya dapat menggunakan fitur cek sentimen untuk melihat hasil klasifikasi. Sementara itu, Admin memiliki akses penuh terhadap semua modul sistem, seperti autentikasi, pengelolaan dashboard, dataset, preprocessing, perhitungan *Naïve Bayes*, dan manajemen pengguna. Setiap fitur dalam diagram menunjukkan fungsi utama sistem dan peran masing-masing pengguna dalam proses interaksi, dari impor data hingga evaluasi hasil klasifikasi.

Pengumpulan Data

Data dalam penelitian ini diperoleh melalui proses web crawling dari platform media sosial X

Penerapan Metode Naïve Bayes Classifier

Proses penerapan metode *data mining* dimulai dengan tahap *preprocessing* yang terdiri dari *lowercasing*, *remove punctuation*, *tokenizing*, *stopword removal*, dan *stemming*. Setelah data dibersihkan, dilakukan proses pelabelan, di mana sistem secara otomatis memberikan label sentimen awal (positif atau negatif) berdasarkan keberadaan kata-kata bernada positif atau negatif dalam teks. Selanjutnya, dilakukan augmentasi untuk menyeimbangkan jumlah data positif dan negatif. Augmentasi dilakukan dengan menyisipkan dan menukar kata dalam kalimat untuk memperkaya variasi data latih. Setelah itu, bobot setiap kata dihitung menggunakan metode *TF-IDF* yang berfungsi sebagai fitur dalam proses klasifikasi sentimen. Dataset awal terdiri dari 1.002 data, dan setelah augmentasi jumlah data meningkat menjadi 1.324 entri. Data ini kemudian dibagi menjadi 75% data latih dan 25% data uji. Proses klasifikasi dilakukan dengan algoritma *Naïve Bayes* untuk mengidentifikasi sentimen positif dan negatif.

Implementasi Sistem



Gambar 4 Implementasi Sistem

Gambar 4 merupakan implementasi sistem yang menampilkan hasil analisis sentimen publik terhadap perpindahan ibu kota ke IKN secara visual dan informatif. Pengguna dapat melihat distribusi sentimen positif dan negatif melalui diagram lingkaran serta kata-kata yang sering

muncul dalam bentuk *Wordcloud*. Selain itu, sistem menyediakan menu navigasi untuk mengelola Dataset, melakukan Preprocessing data, hingga menjalankan klasifikasi menggunakan Naïve Bayes Classifier. Tidak hanya itu, sistem juga menyediakan menu Users yang memungkinkan admin untuk menambahkan, mengubah, menghapus, dan melihat data pengguna sistem.

Hasil Pengujian

Prediksi dilakukan terhadap data uji dan disajikan dalam bentuk tabel Tabel 2.

Tabel 2 Hasil Pengujian

Nilai Prediksi	Nilai Aktual	
	Positive	Negative
Positive	163	3
Negative	22	144

Prediksi dilakukan terhadap data uji dan divisualisasikan dalam bentuk Confusion Matrix. Berdasarkan hasil pengujian, terdapat 144 data negatif yang diklasifikasikan dengan benar (True Negative), 163 data positif diklasifikasikan dengan benar (True Positive), 22 data negatif salah diklasifikasikan sebagai positif (False Positive), dan 3 data positif salah diklasifikasikan sebagai negatif (False Negative).

Evaluasi Kinerja Model

Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan F1-score yang dirangkum dalam Tabel 3.

Tabel 3 Evaluasi Kinerja Model

	Precision	Recall	F1-Score
Negatif	97,96%	86,75%	92,01%
Positif	88,11%	98,19%	92,88%
Macro Average	93,03%	92,47%	92,45%
Accuracy 92,47%			

Evaluasi dilakukan berdasarkan metrik akurasi, presisi, recall, dan F1-score. Akurasi model mencapai 92,47%. Untuk sentimen negatif, presisi sebesar 97,96%, recall sebesar 86,75%, dan F1-score sebesar 92,01%. Untuk sentimen positif, presisi sebesar 88,11%, recall sebesar 98,19%, dan F1-score sebesar 92,88%. Nilai macro average presisi sebesar 93,03%, recall 92,47%, dan F1-score 92,45%.

Pembahasan

Hasil penelitian menunjukkan bahwa performa model yang dikembangkan lebih unggul dibandingkan dengan penelitian sebelumnya. Penelitian sebelumnya hanya memperoleh akurasi sebesar 60,61%, 66%, dan dalam kisaran 77% hingga 86% [12][13][14][15][16]. Keunggulan penelitian ini terletak pada tahapan preprocessing

yang lebih lengkap, penerapan augmentasi, dan penggunaan TF-IDF sebagai bobot fitur.

Pengujian dilakukan dengan pembagian data secara proporsional serta visualisasi Confusion Matrix yang memudahkan interpretasi hasil klasifikasi. Penelitian ini juga memberikan kontribusi signifikan dalam pengolahan opini publik terhadap kebijakan strategis, dengan pendekatan analitik berbasis data dan visualisasi hasil yang informatif.

4. PENUTUP

Kesimpulan

Kesimpulan dari penelitian ini mengacu pada tujuan yang telah ditetapkan, yaitu untuk menganalisis sentimen publik terhadap perpindahan Ibu Kota Negara (IKN) menggunakan algoritma *Naïve Bayes Classifier*. Penelitian berhasil menerapkan metode tersebut terhadap 1.002 data tweet yang diperoleh melalui proses *web crawling* dari media sosial X (sebelumnya Twitter), dengan menggunakan kata kunci “ikn”. Data diproses melalui tahapan *preprocessing* seperti *case folding*, penghapusan tanda baca, tokenisasi, *stopword removal*, *stemming*, serta pembentukan fitur dengan metode TF-IDF. Selanjutnya, data diperkuat melalui proses augmentasi hingga menjadi 1.324 data, lalu dibagi ke dalam data latih dan data uji dengan proporsi 75% : 25%. Klasifikasi sentimen dibatasi pada dua kategori, yaitu positif dan negatif.

Hasil evaluasi performa model menggunakan *Confusion Matrix* menunjukkan akurasi sebesar 92,47%, presisi 93,03%, *recall* 92,47%, dan *F1-score* sebesar 92,45%. Sementara itu, pengujian fungsionalitas sistem dilakukan melalui metode *black-box* dan menunjukkan bahwa sistem dapat berjalan sesuai dengan fungsinya. Berdasarkan hasil tersebut, dapat disimpulkan bahwa sistem berhasil memenuhi kedua tujuan penelitian, yaitu mengimplementasikan algoritma klasifikasi dengan akurasi tinggi dan menghasilkan sistem yang berjalan secara optimal.

Saran

Berdasarkan temuan yang diperoleh dalam penelitian ini, terdapat beberapa saran untuk pengembangan lebih lanjut. Pertama, karena klasifikasi sentimen pada penelitian ini hanya mencakup dua kelas, yaitu positif dan negatif, disarankan untuk menambahkan kelas netral agar hasil analisis lebih komprehensif dan representatif terhadap opini publik. Kedua, proses *crawling* yang saat ini dilakukan secara terpisah

menggunakan Python dapat diintegrasikan ke dalam sistem, sehingga seluruh proses – mulai dari pengambilan data hingga analisis sentimen – dapat berjalan secara otomatis dalam satu sistem terpusat. Ketiga, pengembangan sistem ke depan dapat mencakup fitur penyesuaian kata kunci (*custom keyword*) agar pengguna dapat menentukan topik atau isu tertentu yang ingin dianalisis secara fleksibel dan sesuai kebutuhan.

5. REFERENSI

- [1] O. P. Zusrotun, A. C. Murti, and R. Fiati, “Analisis Sentimen Terhadap Belajar Online pada Media Sosial Twitter Menggunakan Algoritma Naive Bayes,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 310–319, Dec. 2022, doi: 10.23887/janapati.v11i3.49160.
- [2] M. K. Anam, B. N. Pikir, and M. B. Firdaus, “Penerapan Naive Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemerintah,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 139–150, Nov. 2021, doi: 10.30812/matrik.v21i1.1092.
- [3] A. Muzaki and A. Witanti, “Sentiment Analysis Of The Community In The Twitter To The 2020 Election In Pandemic Covid-19 By Method Naive Bayes Classifier,” *Jurnal Teknik Informatika (Jutif)*, vol. 2, no. 2, pp. 101–107, Mar. 2021, doi: 10.20884/1.jutif.2021.2.2.51.
- [4] S. Mulya, H. Sujaini, J. H. Hadari Nawawi, and K. Pontianak, “Analisis Sentimen Tren Olahraga di Masa Pandemi COVID-19 pada Twitter dengan Metode Naive Bayes Classifier (NBC),” *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 2, pp. 284–291, 2022.
- [5] L. E. Pradana and Y. Ruldeviyani, “Sentiment Analysis of Nanovest Investment Application Using Naive Bayes Algorithm,” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 12, no. 2, pp. 283–293, Jul. 2023, doi: 10.23887/janapati.v12i2.62302.
- [6] W. Purbaratri, H. D. Purnomo, D. Manongga, I. Setyawan, and H. Hendry, “Sentiment Analysis of e-Government Service Using the Naive Bayes Algorithm,” *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 441–452, Mar. 2024, doi: 10.30812/matrik.v23i2.3272.
- [7] D. Purnamasari *et al.*, *Pengantar Metode Analisis Sentimen*, 1st ed. Jawa Barat: Gunadarma Penerbit, 2023.
- [8] Okfalisa, *TF-IDF dan Naive Bayes Untuk Filtering dan Monitoring Diskusi Online*, 1st ed. Riau: Al-Mujtahadah Press, 2016.
- [9] Mustika *et al.*, *Data Mining dan Aplikasinya*. Jawa Barat: Widina Bhakti Persada Bandung, 2021. [Online]. Available: www.penerbitwidina.com
- [10] J. Han, J. Pei, and H. Tong, *Data Mining Concepts and Techniques*, 4th ed. United States: Morgan Kaufmann Publishers, 2023.
- [11] A. A. Permana *et al.*, *Machine Learning*, 1st ed. Padang: Global Eksekutif Teknologi, 2023. [Online]. Available: www.globaleksekutifteknologi.co.id
- [12] A. Kurniawan and S. Waluyo, “Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter,” *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, Sep. 2022, [Online]. Available: <https://senafti.budiluhur.ac.id/index.php>
- [13] F. Zamzami, R. Hidayat, and R. Fathonah, “Penerapan Algoritma Naive Bayes Classifier untuk Analisis Sentimen Komentar Twitter Proyek Pembangunan IKN,” *Faktor Exacta*, vol. 17, no. 1, pp. 45–57, May 2024, doi: 10.30998/faktorexacta.v17i1.22265.
- [14] J. Dermawan, Yusra, M. Fikry, S. Agustian, and L. Oktavia, “Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 3, pp. 600–610, Mar. 2024.
- [15] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN,” *Jurnal KomtekInfo*, vol. 10, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [16] R. A. Hakim, W. Gata, Z. P. A. Widodo, O. Kurniawan, and R. A. Syarif, “Analisis Perbandingan Algoritma Machine Learning Terhadap Sentimen Analisis Pemindahan Ibu Kota Negara,” *Jurnal JTIIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 2, Feb. 2023, doi: 10.35870/jti.