



Analisis Sentimen Publik Terhadap Progres Pembangunan IKN di TikTok Menggunakan Naïve Bayes dan SVM

Candra Naya*, Ermanto

Fakultas Teknik, Program Studi Teknik Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}candranaya@pelitabangsa.ac.id, ²ermanto@pelitabangsa.ac.id

Email Penulis Korespondensi: candranaya@pelitabangsa.ac.id

Abstrak—Perkembangan media sosial telah membuka ruang baru bagi masyarakat untuk mengekspresikan opini terhadap berbagai kebijakan publik, termasuk pembangunan Ibu Kota Nusantara (IKN). TikTok sebagai salah satu platform dengan tingkat interaksi tinggi menjadi sumber data yang relevan untuk menganalisis persepsi publik terhadap proyek pembangunan nasional tersebut. Penelitian ini bertujuan untuk menganalisis sentimen komentar TikTok terkait progres pembangunan IKN serta membandingkan performa algoritma klasifikasi Naïve Bayes dan Support Vector Machine (SVM). Penelitian menggunakan pendekatan kuantitatif berbasis data mining dengan kerangka kerja SEMMA yang meliputi tahapan Sample, Explore, Modify, Model, dan Assess. Dataset diperoleh melalui teknik web scraping menggunakan Apify dengan total 2.000 komentar, yang setelah proses seleksi menghasilkan 1.850 komentar valid. Proses preprocessing meliputi cleaning, case folding, tokenizing, stopword removal, dan filtering, kemudian dilakukan ekstraksi fitur menggunakan metode TF-IDF. Dataset dibagi menjadi data latih dan data uji dengan rasio 80:20. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma SVM memiliki performa lebih baik dengan akurasi sebesar 91,25%, precision 90,70%, recall 92,86%, dan F1-score 91,77%, sedangkan Naïve Bayes memperoleh akurasi 84,25%, precision 83,87%, recall 86,67%, dan F1-score 85,24%. Hasil analisis sentimen menunjukkan bahwa sentimen positif terhadap pembangunan IKN sedikit lebih dominan dibandingkan sentimen negatif. Temuan ini menunjukkan bahwa algoritma SVM lebih efektif dalam mengklasifikasikan sentimen komentar media sosial yang memiliki karakteristik bahasa informal seperti pada platform TikTok.

Kata kunci: Analisis Sentimen; TikTok; IKN; Naïve Bayes; Support Vector Machine; Data Mining

Abstract—The rapid development of social media has created new spaces for the public to express opinions regarding various public policies, including the development of Indonesia's new capital city, Nusantara (IKN). TikTok, as one of the platforms with high user interaction, provides a valuable data source for analyzing public perceptions of this national development project. This study aims to analyze sentiment in TikTok comments related to the progress of IKN development and to compare the performance of the Naïve Bayes and Support Vector Machine (SVM) classification algorithms. The research employs a quantitative approach using a data mining framework based on the SEMMA methodology, which includes the stages of Sample, Explore, Modify, Model, and Assess. The dataset was collected through web scraping using Apify, resulting in 2,000 comments, of which 1,850 valid comments remained after the selection process. Text preprocessing was performed through cleaning, case folding, tokenizing, stopword removal, and filtering, followed by feature extraction using the TF-IDF method. The dataset was divided into training and testing sets using an 80:20 ratio. Model evaluation was conducted using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results show that the SVM algorithm outperformed Naïve Bayes with an accuracy of 91.25%, precision of 90.70%, recall of 92.86%, and F1-score of 91.77%, while Naïve Bayes achieved an accuracy of 84.25%, precision of 83.87%, recall of 86.67%, and F1-score of 85.24%. The sentiment distribution indicates that positive sentiment toward the development of IKN slightly dominates negative sentiment. These findings suggest that SVM is more effective for classifying sentiment in informal social media text such as TikTok comments.

Keywords: Sentiment Analysis; TikTok; IKN; Naïve Bayes; Support Vector Machine; Data Mining

1. PENDAHULUAN

Transformasi digital telah mengubah cara masyarakat berinteraksi, menyampaikan aspirasi, serta membentuk opini terhadap kebijakan publik [1]. Media sosial tidak lagi sekadar ruang berbagi konten hiburan, tetapi telah berkembang menjadi arena diskursus publik yang dinamis dan terbuka [2]. Berbagai isu strategis nasional kini dengan cepat memperoleh respons masyarakat melalui komentar, unggahan, dan interaksi digital. Salah satu isu yang menimbulkan perhatian luas adalah pembangunan Ibu Kota Nusantara (IKN) [3] sebagai proyek strategis nasional yang dirancang untuk mendorong pemerataan pembangunan, mengurangi beban Jakarta, serta membangun pusat pemerintahan baru yang berkonsep kota pintar dan berkelanjutan [4]. Proyek ini memunculkan spektrum opini yang luas, mulai dari dukungan optimistis terhadap masa depan pembangunan nasional hingga kritik terkait pembiayaan, dampak lingkungan, kesiapan infrastruktur, dan urgensi kebijakan. Permasalahan yang muncul adalah bagaimana mengidentifikasi, mengukur, dan memetakan sentimen publik tersebut secara sistematis di tengah derasnya arus informasi digital yang tidak terstruktur. Di antara berbagai platform media sosial, TikTok menjadi salah satu medium dengan pertumbuhan pengguna yang sangat signifikan di Indonesia [5]. Karakteristik TikTok yang berbasis video pendek dengan kolom komentar aktif menciptakan ruang interaksi yang ekspresif dan spontan [6]. Komentar pengguna sering kali mengandung opini langsung, respons emosional, serta penggunaan bahasa informal, singkatan, bahkan bahasa gaul yang khas generasi digital [7]. Volume komentar yang besar serta struktur bahasa yang tidak baku menjadikan analisis manual tidak lagi efektif. Kondisi ini menuntut pendekatan komputasional berbasis machine learning yang mampu mengolah data teks dalam jumlah besar secara cepat dan akurat. Dalam konteks ini, analisis sentimen menjadi pendekatan yang relevan untuk mengekstraksi opini publik dan mengelompokkannya ke dalam kategori tertentu



sehingga dapat memberikan gambaran yang terukur mengenai persepsi masyarakat terhadap progres pembangunan IKN[8].

Dalam beberapa tahun terakhir, penelitian mengenai analisis sentimen pada media sosial menunjukkan perkembangan yang signifikan, khususnya dalam penerapan berbagai algoritma klasifikasi. Sebuah penelitian yang dipublikasikan pada tahun 2024 dalam jurnal Buana Informatika menganalisis sentimen terhadap TikTok Shop menggunakan metode Naïve Bayes, K-Nearest Neighbor, dan Decision Tree [9], dan menemukan bahwa algoritma Naïve Bayes menghasilkan tingkat akurasi tertinggi. Hasil penelitian tersebut menunjukkan bahwa Naïve Bayes memiliki performa yang baik dalam menangani teks media sosial, namun penelitian tersebut tidak membahas isu kebijakan publik maupun melakukan perbandingan secara khusus dengan algoritma Support Vector Machine dalam konteks pembangunan nasional. Penelitian lain membandingkan algoritma Naïve Bayes dan Support Vector Machine untuk klasifikasi berita hoaks daring di Indonesia, dan menunjukkan bahwa Support Vector Machine memiliki performa yang lebih baik dibandingkan Naïve Bayes dalam klasifikasi teks berita [10]. Meskipun demikian, kajian tersebut berfokus pada konten berita daring yang memiliki struktur bahasa relatif formal, sehingga karakteristik datanya berbeda dengan komentar media sosial yang umumnya bersifat lebih spontan dan tidak baku [11]. Penelitian lain juga melakukan perbandingan kinerja Naïve Bayes dan Support Vector Machine pada analisis sentimen Twitter terkait pembangunan Ibu Kota Nusantara dan menunjukkan bahwa SVM memiliki tingkat akurasi yang lebih tinggi [12]. Namun demikian, sumber data yang digunakan berasal dari Twitter yang cenderung berbasis teks singkat dan relatif lebih formal dibandingkan komentar pada platform TikTok yang lebih ekspresif, variatif, dan sering menggunakan bahasa informal [13]. Selain itu, terdapat pula penelitian mengenai analisis sentimen terhadap pemindahan IKN menggunakan pendekatan oversampling SMOTE dengan perbandingan algoritma Support Vector Machine dan Random Forest [14], [15]. Penelitian tersebut menunjukkan bahwa penanganan ketidakseimbangan data dapat meningkatkan performa model klasifikasi. Namun penelitian tersebut lebih menitikberatkan pada teknik penyeimbangan data dan belum secara khusus membandingkan algoritma Naïve Bayes dan Support Vector Machine pada data komentar TikTok yang memiliki karakteristik linguistik berbeda dibandingkan teks berita atau Twitter [16].

Berdasarkan kajian terhadap penelitian-penelitian terdahulu tersebut, terlihat adanya kesenjangan penelitian yang cukup jelas. Sebagian besar penelitian masih berfokus pada platform Twitter sebagai sumber data utama, sementara TikTok sebagai platform dengan tingkat interaksi yang sangat tinggi dan dominasi pengguna generasi muda belum banyak dieksplorasi dalam konteks analisis kebijakan publik. Selain itu, meskipun terdapat beberapa penelitian terkait isu pembangunan IKN [17], penelitian yang secara khusus menganalisis persepsi masyarakat terhadap progres pembangunan IKN berdasarkan komentar TikTok masih sangat terbatas. Perbandingan langsung antara algoritma Naïve Bayes dan Support Vector Machine dalam konteks data TikTok juga masih jarang dilakukan, terutama dengan pendekatan preprocessing yang komprehensif untuk menangani karakteristik bahasa informal yang sering muncul pada komentar media sosial [18]. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan memanfaatkan komentar TikTok sebagai sumber data utama dalam menganalisis sentimen masyarakat terhadap progres pembangunan IKN [19]. Penelitian ini tidak hanya berfokus pada identifikasi sentimen positif dan negatif, tetapi juga membandingkan performa dua algoritma klasifikasi yang populer dalam analisis teks, yaitu Naïve Bayes dan Support Vector Machine [20], pada data komentar berbahasa Indonesia yang bersifat tidak terstruktur. Proses analisis dilakukan melalui tahapan preprocessing yang meliputi pembersihan data (cleaning), tokenisasi, transformasi huruf (case folding), penghapusan stopword, serta proses penyaringan kata (filtering), sehingga kualitas dataset dapat ditingkatkan sebelum dilakukan proses pemodelan.

Evaluasi model dilakukan menggunakan confusion matrix untuk mengukur performa klasifikasi berdasarkan nilai akurasi, precision, dan recall secara objektif [21]. Dengan pendekatan tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap progres pembangunan IKN berdasarkan komentar TikTok, membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine dalam klasifikasi sentimen, serta menentukan algoritma yang paling efektif berdasarkan hasil evaluasi performa model [22].

Kajian ini diharapkan memberikan kontribusi secara teoritis, metodologis, dan praktis. Secara teoritis, kajian ini memperkaya studi analisis sentimen berbasis media sosial dengan memanfaatkan platform TikTok sebagai sumber data opini publik. Secara metodologis, kajian ini memberikan referensi mengenai perbandingan performa algoritma klasifikasi pada data teks informal berbahasa Indonesia. Secara praktis, hasil kajian ini dapat menjadi sumber informasi bagi pemerintah dan pemangku kebijakan dalam memahami persepsi masyarakat terhadap pembangunan Ibu Kota Nusantara secara berbasis data..

2. METODOLOGI PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis data mining untuk menganalisis sentimen komentar pengguna TikTok terkait progres pembangunan Ibu Kota Nusantara (IKN), menggunakan kerangka kerja SEMMA.



3.2 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

3.1.1 Tahap Sample (Pengumpulan Data)

Tahap sample merupakan proses pengumpulan data komentar pengguna dari platform media sosial TikTok yang berkaitan dengan pembangunan Ibu Kota Nusantara (IKN). Data diperoleh menggunakan teknik web scraping melalui layanan Apify dengan kata kunci seperti “IKN”, “Ibu Kota Nusantara”, dan “progres pembangunan IKN”. Data yang dikumpulkan berupa teks komentar publik yang masih dalam bentuk data mentah (raw data) dan belum melalui proses pembersihan maupun pengolahan lebih lanjut.

Tabel 1. Tahap Sample

| Komponen | Keterangan |
|-------------------------|---|
| Sumber Data | Media Sosial TikTok |
| Teknik Pengumpulan | Web Scraping menggunakan Apify |
| Kata Kunci | “IKN”, “Ibu Kota Nusantara”, “Progres Pembangunan IKN” |
| Jenis Data | Teks komentar pengguna TikTok beserta metadata pengguna |
| Jumlah Data Crawling | 2.000 komentar |
| Dataset Setelah Seleksi | 1.850 komentar |
| Output | Dataset mentah berupa komentar, ID pengguna, dan tautan video |

Tabel 1 menjelaskan dataset yang dihasilkan terdiri dari beberapa atribut utama, yaitu teks komentar (text), identitas pengguna (uid dan uniqueId), serta tautan video TikTok tempat komentar dipublikasikan (videoWebUrl). Data tersebut masih dalam bentuk data mentah (raw data) yang selanjutnya diproses pada tahap eksplorasi dan preprocessing sebelum dilakukan analisis sentimen.

3.1.2 Tahap Explore

Tahap explore bertujuan untuk memahami karakteristik dataset yang telah dikumpulkan serta melakukan seleksi awal terhadap data komentar. Pada tahap ini dilakukan proses identifikasi komentar duplikat, penghapusan komentar kosong, serta penyaringan data yang tidak relevan dengan topik penelitian. Tahap eksplorasi ini penting untuk memastikan kualitas dan konsistensi data sebelum memasuki proses preprocessing dan analisis lebih lanjut.

Tabel 2. Tahap Explore

| Aktivitas | Tujuan | Output |
|-----------------------------|-------------------------------|------------------|
| Identifikasi duplikasi | Menghindari bias data | Data unik |
| Penghapusan komentar kosong | Menjamin validitas data | Data valid |
| Seleksi relevansi topik | Memastikan kesesuaian konteks | Dataset terpilih |

Tabel 2 menjelaskan aktivitas eksplorasi data yang dilakukan untuk meningkatkan kualitas dataset, meliputi identifikasi duplikasi data, penghapusan komentar kosong, serta seleksi komentar yang relevan dengan topik penelitian sehingga menghasilkan dataset yang lebih valid dan siap diproses pada tahap berikutnya.

3.1.3 Tahap Modify

Tahap modify merupakan proses text preprocessing yang bertujuan untuk menyiapkan data komentar agar dapat diproses dalam model klasifikasi. Mengingat komentar pada media sosial TikTok umumnya bersifat informal dan mengandung berbagai karakter yang tidak relevan, tahap ini dilakukan untuk membersihkan serta menstandarkan teks. Proses preprocessing yang dilakukan meliputi cleaning, case folding, tokenizing, stopword removal, dan filtering. Tahapan ini menghasilkan dataset yang lebih terstruktur sehingga siap digunakan pada proses ekstraksi fitur dan pemodelan klasifikasi.

Tabel 3. Tahap Modify (Preprocessing)

| Sub-Tahap | Deskripsi | Output |
|------------------|---|------------------------------|
| Cleaning | Menghapus URL, simbol, angka, emoji | Teks bersih |
| Case Folding | Mengubah seluruh teks menjadi huruf kecil | Teks seragam |
| Tokenizing | Memecah kalimat menjadi kata | Token kata |
| Stopword Removal | Menghapus kata umum (dan, yang, di, dll.) | Kata bermakna |
| Filtering | Menghapus kata terlalu pendek/panjang | Dataset final siap ekstraksi |

Tabel 3 menjelaskan tahapan text preprocessing yang dilakukan untuk membersihkan dan menstandarkan data komentar, mulai dari penghapusan karakter yang tidak relevan hingga proses pemisahan kata, sehingga menghasilkan dataset yang siap digunakan dalam tahap analisis sentimen.



3.1.4 Tahap Feature Extraction (TF-IDF)

Tahap *feature extraction* bertujuan untuk mengubah data teks komentar menjadi representasi numerik yang dapat diproses oleh algoritma klasifikasi. Pada penelitian ini digunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) untuk menghitung bobot setiap kata dalam dataset komentar. Metode ini memberikan bobot yang lebih tinggi pada kata yang sering muncul dalam suatu dokumen tetapi jarang muncul pada dokumen lainnya, sehingga kata yang memiliki informasi penting dalam menentukan sentimen dapat direpresentasikan secara lebih signifikan. Hasil dari proses ini berupa matriks fitur numerik yang selanjutnya digunakan sebagai input pada tahap pemodelan klasifikasi.

Tabel 4. Tahap Feature Extraction

| Komponen | Keterangan |
|----------|---|
| Metode | TF-IDF |
| Fungsi | Mengubah teks menjadi data numerik |
| Output | Matriks fitur numerik |
| Manfaat | Meningkatkan kualitas representasi kata |

Tabel 4 menjelaskan metode ekstraksi fitur yang digunakan untuk mengubah dataset komentar berbasis teks menjadi representasi numerik menggunakan teknik TF-IDF. Proses ini menghasilkan matriks fitur yang merepresentasikan bobot setiap kata dalam dokumen sehingga data dapat diproses lebih lanjut pada tahap pemodelan menggunakan algoritma klasifikasi.

Pada tahap ekstraksi fitur, teks komentar yang telah melalui proses preprocessing kemudian diubah menjadi representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dibandingkan dengan keseluruhan dataset. Kata yang sering muncul pada suatu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot lebih tinggi sehingga dianggap lebih representatif dalam menggambarkan isi teks.

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \tag{1}$$

$$IDF(t) = \log \frac{N}{df(t)} \tag{2}$$

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \tag{3}$$

- $f_{t,d}$: frekuensi term t dalam dokumen d
- N : jumlah total dokumen
- $df(t)$: jumlah dokumen yang memuat term t

Pembobotan fitur dilakukan menggunakan metode TF-IDF. Nilai TF merepresentasikan frekuensi kemunculan term dalam dokumen, sedangkan IDF menunjukkan tingkat kepentingan term dalam seluruh koleksi dokumen. Bobot akhir term diperoleh dari hasil perkalian TF dan IDF.

3.1.5 Tahap Model

Tahap model merupakan proses pelatihan dan pengujian algoritma klasifikasi untuk mengidentifikasi sentimen komentar pengguna TikTok terhadap pembangunan Ibu Kota Nusantara. Pada tahap ini digunakan dua algoritma machine learning, yaitu Naïve Bayes dan Support Vector Machine (SVM). Dataset yang telah melalui proses preprocessing dan ekstraksi fitur menggunakan TF-IDF kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20. Model Naïve Bayes bekerja berdasarkan pendekatan probabilistik untuk menghitung kemungkinan suatu dokumen termasuk dalam kelas tertentu, sedangkan SVM bekerja dengan mencari hyperplane optimal yang mampu memisahkan kelas sentimen positif dan negatif secara maksimal. Hasil dari tahap ini berupa model klasifikasi yang digunakan untuk memprediksi sentimen pada dataset komentar.

Tabel 5. Tahap Model

| Komponen | Naïve Bayes | Support Vector Machine |
|------------|-----------------------------|-----------------------------|
| Pendekatan | Probabilistik | Margin-based classifier |
| Kelebihan | Cepat & efisien | Akurasi tinggi |
| Input | Matriks TF-IDF | Matriks TF-IDF |
| Output | Klasifikasi Positif/Negatif | Klasifikasi Positif/Negatif |

Tabel 5 menjelaskan perbandingan karakteristik dua algoritma klasifikasi yang digunakan dalam penelitian, yaitu Naïve Bayes dan Support Vector Machine. Tabel tersebut menunjukkan pendekatan metode, kelebihan masing-masing algoritma, serta input berupa matriks fitur TF-IDF dan output berupa hasil klasifikasi sentimen positif atau negatif.

Algoritma Naïve Bayes merupakan metode klasifikasi berbasis probabilistik yang memanfaatkan Teorema Bayes untuk menentukan probabilitas suatu dokumen termasuk ke dalam kelas tertentu. Metode ini mengasumsikan bahwa setiap fitur bersifat independen terhadap fitur lainnya. Dengan pendekatan ini, probabilitas suatu dokumen teks



termasuk dalam kelas sentimen tertentu dapat dihitung berdasarkan distribusi probabilitas dari fitur-fitur yang muncul dalam dokumen tersebut.

$$P(c | d) = \frac{P(d|c)P(c)}{P(d)} \quad (4)$$

Pada algoritma Naïve Bayes, probabilitas suatu dokumen d termasuk ke dalam kelas c dihitung menggunakan Teorema Bayes dengan mempertimbangkan probabilitas prior kelas dan probabilitas kemunculan fitur pada kelas tersebut.

Selain menggunakan Naïve Bayes, penelitian ini juga memanfaatkan algoritma Support Vector Machine (SVM) untuk melakukan klasifikasi sentimen. SVM bekerja dengan mencari hyperplane optimal yang mampu memisahkan dua kelas data secara maksimal dalam ruang fitur berdimensi tinggi. Hyperplane tersebut dibentuk berdasarkan vektor bobot dan bias yang diperoleh dari proses pelatihan model.

$$f(x) = w \cdot x + b \quad (5)$$

SVM bekerja dengan mencari hyperplane optimal $f(x)$ yang mampu memisahkan dua kelas sentimen secara maksimal dalam ruang fitur berdimensi tinggi.

3.1.6 Tahap Assess

Tahap assess dilakukan untuk mengevaluasi kinerja model klasifikasi yang telah dibangun pada tahap sebelumnya. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur performa algoritma dalam mengklasifikasikan sentimen komentar pengguna TikTok. Beberapa parameter evaluasi yang digunakan meliputi accuracy, precision, recall, dan F1-score. Hasil evaluasi dari masing-masing algoritma kemudian dibandingkan untuk mengetahui model yang memiliki performa terbaik dalam mengklasifikasikan sentimen komentar terkait pembangunan Ibu Kota Nusantara.

Tabel 6. Tahap Assess

| Parameter | Fungsi | Rumus Singkat |
|-----------|---|----------------------------|
| Akurasi | Mengukur ketepatan keseluruhan prediksi | $(TP + TN) / \text{Total}$ |
| Precision | Ketepatan prediksi positif | $TP / (TP + FP)$ |
| Recall | Kemampuan mendeteksi data positif | $TP / (TP + FN)$ |
| F1-Score | Keseimbangan precision & recall | $2PR / (P+R)$ |

Tabel 6 menjelaskan parameter evaluasi yang digunakan dalam penelitian untuk mengukur kinerja model klasifikasi. Parameter tersebut meliputi akurasi untuk mengukur ketepatan keseluruhan prediksi, precision untuk mengukur ketepatan prediksi positif, recall untuk mengukur kemampuan model dalam mendeteksi data positif, serta F1-score yang merupakan kombinasi antara precision dan recall.

Untuk mengevaluasi performa model klasifikasi yang dihasilkan, digunakan beberapa metrik evaluasi yang diperoleh dari confusion matrix, yaitu akurasi, precision, recall, dan F1-score. Metrik tersebut digunakan untuk mengukur tingkat ketepatan model dalam memprediksi kelas sentimen serta keseimbangan antara kemampuan model dalam mendeteksi data positif dan meminimalkan kesalahan prediksi.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

Berdasarkan persamaan tersebut, nilai akurasi digunakan untuk mengukur tingkat ketepatan keseluruhan model dalam melakukan klasifikasi terhadap seluruh data uji. Sementara itu, precision menunjukkan tingkat ketepatan prediksi terhadap data yang diprediksi sebagai kelas positif. Recall menggambarkan kemampuan model dalam mendeteksi seluruh data yang benar-benar termasuk dalam kelas positif. Adapun F1-score merupakan nilai harmonisasi antara precision dan recall yang digunakan untuk menilai keseimbangan performa model. Keempat metrik evaluasi tersebut digunakan untuk membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen komentar TikTok terkait pembangunan Ibu Kota Nusantara.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

3.1.1 Hasil Pengumpulan dan Karakteristik Data

Proses pengumpulan data dilakukan melalui teknik web scraping menggunakan layanan Apify pada platform media sosial TikTok dengan kata kunci "IKN", "Ibu Kota Nusantara", dan "progres pembangunan IKN". Tahap crawling



menghasilkan 2.000 komentar mentah yang berkaitan dengan pembangunan Ibu Kota Nusantara. Selanjutnya dilakukan proses seleksi awal untuk menghapus komentar duplikat, komentar kosong, serta komentar yang tidak relevan dengan topik penelitian. Setelah proses seleksi dilakukan, diperoleh 1.850 komentar valid yang digunakan sebagai dataset penelitian. Dataset tersebut mencerminkan karakteristik komunikasi pada media sosial TikTok yang cenderung menggunakan bahasa informal, singkatan, serta ekspresi emosional. Kondisi ini menunjukkan bahwa data komentar TikTok memiliki kompleksitas linguistik yang tinggi sehingga memerlukan proses preprocessing sebelum dilakukan analisis sentimen.

Tabel 7. Ringkasan Dataset

| Keterangan | Jumlah |
|-------------------------------|--------|
| Total komentar hasil crawling | 2.000 |
| Komentar duplikat | 85 |
| Komentar tidak relevan/kosong | 65 |
| Dataset akhir | 1.850 |

Tabel 7 menunjukkan proses seleksi dataset dari hasil crawling hingga diperoleh dataset final sebanyak 1.850 komentar yang digunakan dalam proses analisis sentimen.

3.1.2 Hasil Preprocessing dan Ekstraksi Fitur

Tahap preprocessing dilakukan untuk membersihkan dan menstandarkan teks komentar sebelum dilakukan analisis menggunakan algoritma machine learning. Proses preprocessing meliputi cleaning, case folding, tokenizing, stopword removal, dan filtering.

Tahap cleaning dilakukan dengan menghapus URL, simbol, angka, emoji, serta karakter khusus yang tidak relevan dengan analisis teks. Selanjutnya dilakukan case folding untuk mengubah seluruh huruf menjadi huruf kecil sehingga menghasilkan bentuk teks yang seragam. Proses tokenizing kemudian memecah kalimat komentar menjadi unit kata yang lebih kecil. Setelah itu dilakukan stopword removal untuk menghapus kata umum yang tidak memiliki makna signifikan dalam analisis sentimen. Tahap terakhir adalah filtering untuk menyaring kata yang terlalu pendek atau tidak relevan.

Setelah preprocessing selesai, data teks kemudian diubah menjadi representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF)..

Tabel 8. Hasil Tahap Preprocessing

| Tahapan | Hasil |
|--------------------|----------------|
| Dataset awal | 1.850 komentar |
| Setelah cleaning | 1.850 komentar |
| Jumlah term unik | 3.200 kata |
| Representasi fitur | Matriks TF-IDF |

Tabel 8 menunjukkan hasil preprocessing dataset yang menghasilkan 3.200 term unik yang kemudian digunakan sebagai fitur dalam model klasifikasi sentimen.

3.1.3 Contoh Hasil Preprocessing Komentar TikTok

Untuk memberikan gambaran mengenai proses pembersihan dan normalisasi teks, ditampilkan beberapa contoh komentar TikTok sebelum dan sesudah melalui tahap preprocessing. Proses ini meliputi cleaning, case folding, tokenizing, serta penghapusan stopword sehingga menghasilkan teks yang lebih terstruktur dan siap digunakan pada tahap ekstraksi fitur menggunakan metode TF-IDF.

Tabel 9. Contoh Hasil Preprocessing Komentar TikTok

| Komentar Asli | Setelah Preprocessing |
|--|--|
| saya yang kerja di ikn merasa pembangunan masih baik saja emang dananya kurang berapa bro kebetulan gw bisa bantu katanya banyak investor asing yg antri | kerja ikn merasa pembangunan baik dana kurang berapa bantu banyak investor asing antri |
| kalo bener di stop ikn gua sumpah nyesel banget dulu milih mending lanjutkan klo mangkrak malah rugi | stop ikn nyesal milih lanjut mangkrak rugi |

Tabel 9 menunjukkan contoh perubahan teks komentar TikTok setelah melalui proses preprocessing. Tahapan tersebut meliputi pembersihan karakter yang tidak relevan, perubahan huruf menjadi bentuk seragam, pemisahan kata, serta penghapusan kata umum yang tidak memiliki makna penting. Proses ini bertujuan untuk meningkatkan kualitas data sebelum dilakukan ekstraksi fitur dan proses klasifikasi sentimen.



3.1.4 Analisis Fitur TF-IDF pada Dataset

Setelah proses preprocessing selesai, data komentar TikTok kemudian diubah menjadi representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dibandingkan dengan keseluruhan dataset. Kata yang memiliki bobot TF-IDF tinggi dianggap lebih penting dalam merepresentasikan informasi yang terkandung dalam teks.

Tabel 10. Kata dengan Bobot TF-IDF Tertinggi pada Dataset

| No | Term | Bobot TF-IDF |
|----|-------------|--------------|
| 1 | ikn | 0.084 |
| 2 | pembangunan | 0.078 |
| 3 | lanjut | 0.072 |
| 4 | proyek | 0.069 |
| 5 | investor | 0.066 |
| 6 | pemerintah | 0.063 |
| 7 | ekonomi | 0.061 |
| 8 | mangkrak | 0.058 |
| 9 | anggaran | 0.055 |
| 10 | dukung | 0.052 |

Tabel 10 menunjukkan beberapa kata yang memiliki bobot TF-IDF tertinggi pada dataset komentar TikTok terkait pembangunan Ibu Kota Nusantara. Nilai bobot yang lebih tinggi menunjukkan bahwa kata tersebut memiliki tingkat kepentingan yang lebih besar dalam merepresentasikan isi dokumen. Kata-kata tersebut menjadi fitur penting yang digunakan oleh algoritma klasifikasi dalam menentukan sentimen komentar.

3.1.5 Statistik Dataset Komentar TikTok

Untuk memberikan gambaran mengenai karakteristik dataset yang digunakan dalam penelitian ini, dilakukan analisis statistik terhadap data komentar TikTok yang telah melalui tahap preprocessing dan ekstraksi fitur. Analisis ini mencakup jumlah dataset, distribusi data latih dan data uji, serta jumlah term unik yang dihasilkan dari proses ekstraksi fitur TF-IDF.

Tabel 11. Hasil Tahap Preprocessing

| Statistik Dataset | Nilai |
|-------------------------------|-------|
| Total komentar hasil crawling | 2.000 |
| Dataset setelah seleksi | 1.850 |
| Data latih (80%) | 1.480 |
| Data uji (20%) | 400 |
| Jumlah term unik | 3.200 |
| Sentimen positif (data uji) | 210 |
| Sentimen negatif (data uji) | 190 |

Tabel 11 menunjukkan ringkasan statistik dataset yang digunakan dalam penelitian. Dataset terdiri dari 2.000 komentar hasil proses crawling yang kemudian melalui tahap seleksi sehingga diperoleh 1.850 komentar valid. Dataset tersebut selanjutnya dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% untuk proses pelatihan dan pengujian model klasifikasi.

3.1.6 Hasil Klasifikasi Menggunakan Naïve Bayes

Model Naïve Bayes dilatih menggunakan data latih sebesar 80% dari total dataset dan diuji menggunakan 20% data uji. Pengujian dilakukan menggunakan 400 data uji (20% dari dataset). Hasil klasifikasi Naïve Bayes ditunjukkan dalam confusion matrix berikut:

Tabel 12. Confusion Matrix Naïve Bayes

| | Prediksi Positif | Prediksi Negatif |
|----------------|------------------|------------------|
| Aktual Positif | 182 (TP) | 28 (FN) |
| Aktual Negatif | 35 (FP) | 155 (TN) |

Tabel 12 menunjukkan distribusi hasil prediksi model Naïve Bayes terhadap data uji. Model berhasil mengklasifikasikan 182 komentar positif secara benar, namun masih terdapat kesalahan klasifikasi pada beberapa komentar yang memiliki struktur ambigu atau mengandung makna ganda.

Tabel 13. Evaluasi Performa Naïve Bayes

| Parameter | Nilai |
|-----------|--------|
| Akurasi | 84.25% |



| Parameter | Nilai |
|-----------|--------|
| Precision | 83.87% |
| Recall | 86.67% |
| F1-Score | 85.24% |

Tabel 13 menunjukkan bahwa model Naïve Bayes mampu mencapai akurasi sebesar 84,25%, yang menunjukkan kemampuan klasifikasi yang cukup baik dalam mengidentifikasi sentimen komentar TikTok.

3.1.7 Hasil Klasifikasi Menggunakan Support Vector Machine (SVM)

Algoritma kedua yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM). Model SVM dilatih menggunakan dataset yang sama dengan skema pembagian data yang identik dengan model Naïve Bayes.

Tabel 14. Confusion Matrix SVM

| | Prediksi Positif | Prediksi Negatif |
|----------------|------------------|------------------|
| Aktual Positif | 195 (TP) | 15 (FN) |
| Aktual Negatif | 20 (FP) | 170 (TN) |

Tabel 14 menunjukkan bahwa model SVM mampu mengklasifikasikan lebih banyak komentar secara benar dibandingkan Naïve Bayes, dengan jumlah kesalahan prediksi yang lebih kecil.

Tabel 15. Evaluasi Performa SVM

| Parameter | Nilai |
|-----------|--------|
| Akurasi | 91.25% |
| Precision | 90.70% |
| Recall | 92.86% |
| F1-Score | 91.77% |

Tabel 15 menunjukkan bahwa model SVM menghasilkan performa yang lebih tinggi pada seluruh metrik evaluasi dibandingkan Naïve Bayes.

3.1.8 Perbandingan Performa Model

Untuk melihat perbedaan performa kedua algoritma secara komprehensif, dilakukan perbandingan berdasarkan nilai akurasi, precision, recall, dan F1-score.

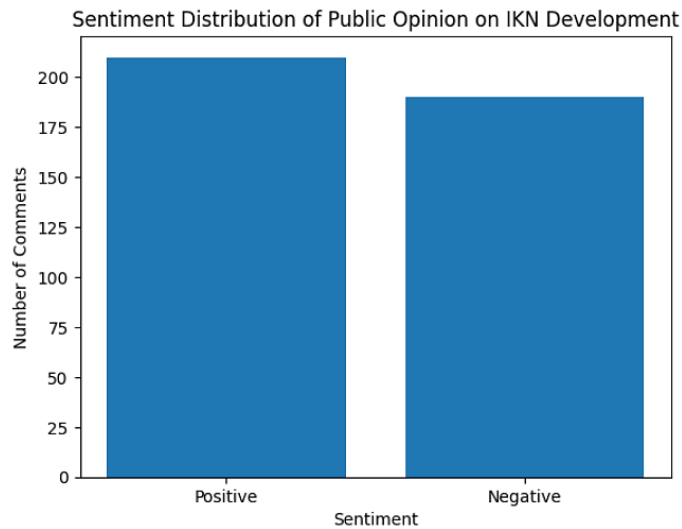
Tabel 16. Perbandingan Performa Model

| Parameter | Naïve Bayes | SVM |
|-----------|-------------|--------|
| Akurasi | 84.25% | 91.25% |
| Precision | 83.87% | 90.70% |
| Recall | 86.67% | 92.86% |
| F1-Score | 85.24% | 91.77% |

Tabel 16 menunjukkan bahwa Support Vector Machine memiliki performa lebih tinggi pada seluruh metrik evaluasi, sehingga algoritma ini lebih efektif dalam mengklasifikasikan sentimen komentar TikTok.

3.2 Pembahasan

Hasil analisis menunjukkan bahwa komentar TikTok dapat dimanfaatkan sebagai sumber data yang relevan untuk memahami opini publik terhadap pembangunan Ibu Kota Nusantara. Distribusi sentimen pada data uji menunjukkan bahwa sentimen positif dan negatif relatif berimbang, meskipun sentimen positif sedikit lebih dominan. Sentimen positif umumnya berkaitan dengan harapan terhadap pembangunan infrastruktur dan pertumbuhan ekonomi, sedangkan sentimen negatif lebih banyak berkaitan dengan kekhawatiran mengenai pembiayaan proyek dan dampak lingkungan. Dari sisi performa algoritma, Support Vector Machine menunjukkan kinerja yang lebih baik dibandingkan Naïve Bayes dalam mengklasifikasikan sentimen komentar TikTok. Hal ini menunjukkan bahwa SVM lebih efektif dalam menangani data teks berdimensi tinggi hasil pembobotan TF-IDF. Meskipun demikian, Naïve Bayes tetap menunjukkan performa yang cukup baik dengan akurasi di atas 80%. Oleh karena itu, SVM dapat direkomendasikan sebagai metode yang lebih optimal untuk analisis sentimen pada data media sosial yang memiliki karakteristik bahasa informal seperti komentar TikTok.



Gambar 2. Distribusi Sentimen Publik terhadap Progres Pembangunan IKN

Gambar 2 menunjukkan distribusi sentimen publik terhadap pembangunan Ibu Kota Nusantara (IKN) berdasarkan hasil klasifikasi komentar TikTok. Dari total 400 data uji, sebanyak 210 komentar (52,5%) diklasifikasikan sebagai sentimen positif, sedangkan 190 komentar (47,5%) termasuk sentimen negatif. Hasil ini menunjukkan bahwa persepsi masyarakat terhadap pembangunan IKN relatif berimbang, meskipun sentimen positif sedikit lebih dominan. Distribusi ini mencerminkan adanya dukungan masyarakat terhadap pembangunan IKN, namun tetap disertai berbagai pandangan kritis terhadap kebijakan tersebut.



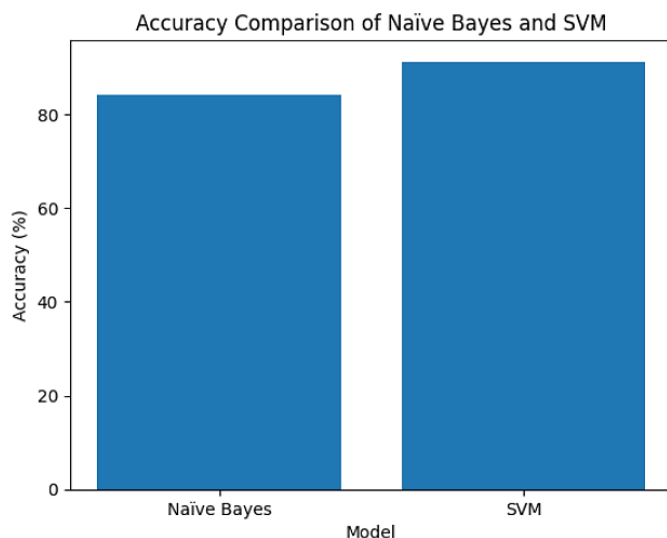
Gambar 3. Word Cloud Sentimen Positif

Gambar 3 menunjukkan kata-kata yang dominan muncul dalam komentar yang mendukung pembangunan Ibu Kota Nusantara. Kata seperti *pembangunan*, *lanjut*, *investor*, dan *ekonomi* menunjukkan adanya harapan masyarakat terhadap pertumbuhan ekonomi serta percepatan pembangunan infrastruktur.



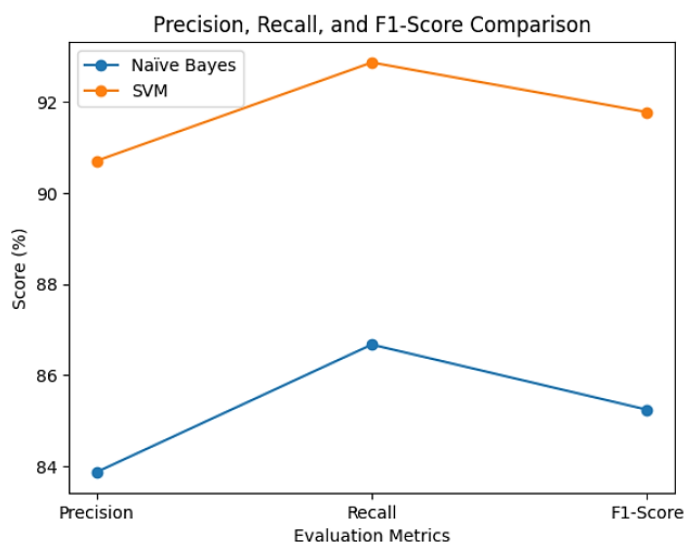
Gambar 4. Word Cloud Sentimen Negatif

Gambar 4 memperlihatkan kata-kata yang sering muncul dalam komentar yang mengandung kritik terhadap pembangunan IKN. Kata seperti *mangkrak*, *anggaran*, *mahal*, dan *risiko* menunjukkan kekhawatiran masyarakat terkait pembiayaan proyek dan dampak pembangunan.



Gambar 5. Perbandingan Akurasi Model Naïve Bayes dan Support Vector Machine

Gambar 5 menunjukkan perbandingan tingkat akurasi antara algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen komentar TikTok terkait pembangunan Ibu Kota Nusantara. Hasil evaluasi menunjukkan bahwa SVM memiliki akurasi lebih tinggi sebesar 91,25% dibandingkan Naïve Bayes yang mencapai 84,25%, sehingga SVM dinilai lebih efektif dalam mengklasifikasikan sentimen pada dataset penelitian ini..



Gambar 6. Perbandingan nilai precision, recall, dan F1-score antara Naïve Bayes dan Support Vector

Gambar 6 menunjukkan perbandingan nilai precision, recall, dan F1-score antara algoritma Naïve Bayes dan Support Vector Machine (SVM) dalam klasifikasi sentimen komentar TikTok terkait pembangunan Ibu Kota Nusantara. Hasil evaluasi menunjukkan bahwa SVM memiliki performa lebih tinggi pada seluruh metrik, dengan nilai precision 90,70%, recall 92,86%, dan F1-score 91,77%, dibandingkan Naïve Bayes yang memperoleh precision 83,87%, recall 86,67%, dan F1-score 85,24%. Hal ini menunjukkan bahwa SVM lebih efektif dalam menghasilkan prediksi sentimen yang akurat dan konsisten pada dataset penelitian.

4. KESIMPULAN

Analisis sentimen terhadap progres pembangunan Ibu Kota Nusantara (IKN) berdasarkan komentar pengguna TikTok menunjukkan bahwa data media sosial dapat dimanfaatkan sebagai sumber informasi yang relevan untuk memahami persepsi masyarakat terhadap kebijakan pembangunan nasional. Proses analisis dilakukan menggunakan pendekatan data mining dengan kerangka kerja SEMMA yang meliputi tahapan pengumpulan data, eksplorasi data, preprocessing teks, ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), pemodelan klasifikasi, serta evaluasi performa model. Hasil analisis menunjukkan bahwa opini publik terhadap pembangunan IKN relatif berimbang antara sentimen positif dan negatif, meskipun sentimen positif sedikit lebih dominan. Sentimen positif umumnya berkaitan dengan harapan masyarakat terhadap pemerataan pembangunan, percepatan pembangunan infrastruktur, serta potensi pertumbuhan ekonomi di kawasan baru. Sebaliknya, sentimen negatif lebih banyak berkaitan



dengan kekhawatiran mengenai pembiayaan proyek, dampak lingkungan, serta prioritas pembangunan nasional. Berdasarkan hasil evaluasi model klasifikasi, algoritma Support Vector Machine menunjukkan performa yang lebih baik dibandingkan Naïve Bayes pada seluruh metrik evaluasi yang digunakan. Model Support Vector Machine memperoleh akurasi sebesar 91,25%, precision sebesar 90,70%, recall sebesar 92,86%, dan F1-score sebesar 91,77%, sedangkan Naïve Bayes memperoleh akurasi sebesar 84,25%, precision sebesar 83,87%, recall sebesar 86,67%, dan F1-score sebesar 85,24%. Hasil ini menunjukkan bahwa Support Vector Machine lebih efektif dalam menangani data teks berdimensi tinggi yang dihasilkan dari pembobotan TF-IDF serta mampu memodelkan kompleksitas hubungan antar kata pada komentar media sosial yang bersifat informal. Beberapa keterbatasan masih terdapat dalam kajian ini, antara lain jumlah dataset yang relatif terbatas serta sumber data yang hanya berasal dari satu platform media sosial, yaitu TikTok. Oleh karena itu, kajian selanjutnya disarankan menggunakan dataset yang lebih besar dan beragam dari berbagai platform media sosial agar diperoleh gambaran opini publik yang lebih komprehensif. Selain itu, pengembangan metode analisis dengan membandingkan algoritma klasifikasi lain seperti Random Forest, Logistic Regression, atau pendekatan deep learning juga dapat dilakukan untuk meningkatkan akurasi serta kemampuan generalisasi model dalam menganalisis sentimen publik terhadap kebijakan pembangunan nasional. Selain itu, pemanfaatan analisis sentimen berbasis data mining pada data media sosial dapat memberikan gambaran yang lebih cepat dan objektif mengenai dinamika opini publik terhadap kebijakan pembangunan nasional. Informasi yang dihasilkan dari analisis tersebut dapat digunakan sebagai bahan pertimbangan bagi pemerintah dan pemangku kepentingan dalam merumuskan strategi komunikasi publik yang lebih efektif. Dengan demikian, pemanfaatan teknologi analisis teks tidak hanya berfungsi sebagai alat analisis akademik, tetapi juga dapat menjadi instrumen pendukung pengambilan keputusan yang berbasis data dalam merespons aspirasi masyarakat di era digital.

REFERENCES

- [1] Duvalio Adnan Zordi, Mohammad Syahrul Ihsan, Muhamad Aprian Nazarudin, and Tria Patrianti, "Peran Media Sosial dalam Pembentukan Opini Publik dan Dinamika Kebijakan Pemerintah di Era Digital," *Jurnal Ilmu Komunikasi, Administrasi Publik dan Kebijakan Negara*, vol. 3, no. 1, pp. 154–162, Jan. 2026, doi: 10.62383/komunikasi.v3i1.882.
- [2] L. A. Andini, "Media Sosial sebagai Ruang Digital Activism Generasi Muda Dalam Memperjuangkan Isu Lingkungan," *Ganaya : Jurnal Ilmu Sosial dan Humaniora*, vol. 8, no. 4, pp. 127–137, Oct. 2025, doi: 10.37329/ganaya.v8i4.4937.
- [3] S. A. S. Mola, Iqbal Muhammad Iskandar, J. E. Pidu Dimu, and W. Y. Seran, "Analisis Sentimen Pembangunan Ibu Kota Negara Indonesia Menggunakan Metode Naïve Bayes, dan K-Nearest Neighbor," *HOAQ (High Education of Organization Archive Quality) : Jurnal Teknologi Informasi*, vol. 15, no. 2, pp. 151–157, Dec. 2024, doi: 10.52972/hoaq.vol15no2.p151-157.
- [4] B. O. Lubis *et al.*, "Analisis Sentimen Publik terhadap Pembangunan Ibu Kota Nusantara (IKN) menggunakan Algoritma Naive Bayes," *Jurnal Ilmiah Sistem Informasi*, vol. 5, no. 1, pp. 502–513, Jan. 2026, doi: 10.51903/nx1aza77.
- [5] S. V. Mahardhika, I. Nurjannah, I. I. Ma'una, and Z. Islamiyah, "Faktor-Faktor Penyebab Tingginya Minat Generasi Post-Millennial Di Indonesia Terhadap Penggunaan Aplikasi Tik-Tok," *SOSEARCH : Social Science Educational Research*, vol. 2, no. 1, pp. 40–53, Dec. 2021, doi: 10.26740/sosearch.v2n1.p40-53.
- [6] Baharudin Ihsan, A. Anzilna Munzalan. M, Laksa Rizal Putra. W, David Yusuf. A, M. Helmi, and Nurudin Nurudin, "Strategi Efektif Penggunaan Media Sosial (TIKTOK) dalam Pembelajaran Agama di Era Digital," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 4, pp. 52–64, Jun. 2024, doi: 10.61132/merkurius.v2i4.139.
- [7] D. A. P. Arnetta and C. G. Haryono, "Pemanfaatan media sosial TikTok sebagai platform pemasaran digital pada akun @dododots.by.zen," *AKADEMIK: Jurnal Mahasiswa Humanis*, vol. 5, no. 2, pp. 1095–1107, May 2025, doi: 10.37481/jmh.v5i2.1451.
- [8] Siti Rihastuti and Afnan Rosyidi, "Analisis Sentimen Pengguna Tiktok tentang Progres Pembangunan IKN dengan Metode Random Forest," *Journal of Computer Science and Technology (JCS-TECH)*, vol. 5, no. 1, pp. 19–23, May 2025, doi: 10.54840/jcstech.v5i1.345.
- [9] F. Fieryando and B. Kristianto, "Analisis Sentimen Terhadap TikTok Shop Dengan K-Nearest Neighbor, Decision Tree, dan Naive Bayes," *Jurnal Buana Informatika*, vol. 15, no. 01, pp. 21–29, Apr. 2024, doi: 10.24002/jbi.v15i1.8205.
- [10] N. Hadi and D. Sugiarto, "Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, Jan. 2025, doi: 10.30591/jpit.v10i1.7106.
- [11] A. Setiawan and R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 183–192, Jun. 2024, doi: 10.29408/edumatic.v8i1.25667.
- [12] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, and R. Rahmaddeni, "Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara," *JURNAL ILMIAH INFORMATIKA*, vol. 12, no. 01, pp. 15–21, Mar. 2024, doi: 10.33884/jif.v12i01.8721.
- [13] D. Novita, A. Herwanto, E. Cahyo Maynardarto, M. Anton Maulana, and H. Hanifah, "Penggunaan Media Sosial TikTok Sebagai Media Promosi Pemasaran Dalam Bisnis Online," *Jurnal Minfo Polgan*, vol. 12, no. 2, pp. 2543–2550, Dec. 2023, doi: 10.33395/jmp.v12i2.13312.
- [14] Ahmad Almas'ud ZD, Kusriani, and Anggit Dwi Hartanto, "Analisis Sentimen Pemindahan Ibu Kota Negara (IKN) Menggunakan Metode Oversampling Synthetic Minority (SMOTE)," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 9, no. 2, pp. 324–335, Dec. 2024, doi: 10.24252/instek.v9i2.50944.
- [15] D. Nurmalasari, H. R. Yuliantoro, and D. H. Qudsi, "Improving Panic Disorder Classification Using SMOTE and Random Forest," *Journal of Applied Informatics and Computing*, vol. 8, no. 2, pp. 272–279, Oct. 2024, doi: 10.30871/jaic.v8i2.8315.
- [16] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/teknov10i2.419.



- [17] N. Hadi and D. Sugiarto, “Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, Jan. 2025, doi: 10.30591/jpit.v10i1.7106.
- [18] D. N. Novianti, D. F. Shiddieq, F. F. Roji, and W. Susilawati, “Komparasi Algoritma Support Vector Machine dan Naïve Bayes untuk Analisis Sentimen pada Metaverse,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 231–239, Jan. 2024, doi: 10.57152/malcom.v4i1.1061.
- [19] S. B. Setiawan and A. R. Isnain, “Sentimen Analisis Masyarakat Terhadap Pembangunan IKN Menggunakan Algoritma Lexicon Based Approach dan Naïve Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 2, p. 1019, Apr. 2024, doi: 10.30865/mib.v8i2.7506.
- [20] M. I. Fikri, T. S. Sabrila, and Y. Azhar, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *SMATIKA JURNAL*, vol. 10, no. 02, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455.
- [21] M. K. Suryadewiansyah and T. E. E. Tju, “Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 8, no. 2, pp. 81–88, Aug. 2022, doi: 10.25077/TEKNOSI.v8i2.2022.81-88.
- [22] R. B. Dahlian and D. Sitanggang, “Sentiment Analysis of Digital Television Migration on Twitter Using Naïve Bayes Multinomial Comparison, Support Vector Machines, and Logistic Regression Algorithms,” *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 2, pp. 280–288, Jul. 2023, doi: 10.32736/sisfokom.v12i2.1668.