

HOAQ: JURNAL TEKNOLOGI INFORMASI, Volume 15 Nomor 2, Desember 2024, 151-157  
p-ISSN: 2337-5280, e-ISSN: 26207427  
DOI: <https://doi.org/10.52972/hoaq.vol15no2.p151-157>

## ANALISIS SENTIMEN PEMBANGUNAN IBU KOTA NEGARA INDONESIA DI YOUTUBE MENGGUNAKAN METODE NAÏVE BAYES, DAN K-NEAREST NEIGHBOR

Sebastianus Adi Santoso Mola<sup>1</sup>, Iqbal Muhammad Iskandar<sup>2</sup>, Joey Elisa Pidu Dimu<sup>3</sup>,  
Willem Yufri Seran<sup>4</sup>

<sup>1,2,3</sup>Program Studi Ilmu Komputer, Universitas Nusa Cendana Kupang  
Jl. Adisucipto Penfui, Kupang – Nusa Tenggara Timur, Indonesia

Email: <sup>1</sup>adimola@staf.undana.ac.id, <sup>2</sup>iqballbarai@gmail.com, <sup>3</sup>joeyelisa16@gmail.com, <sup>4</sup>willemseran1@gmail.com

### ABSTRAK

Pembangunan Ibu Kota Negara (IKN) di Indonesia telah menarik perhatian publik secara luas dan menjadi topik diskusi di media sosial, termasuk platform YouTube. Untuk memahami sentimen publik terkait proyek ini, penelitian ini menggunakan dua metode pembelajaran mesin, yaitu K-Nearest Neighbors (KNN) dan NAÏVE BAYES. Data diperoleh dari komentar-komentar di YouTube yang kemudian diproses melalui beberapa tahapan, termasuk pembersihan data (preprocessing), tokenisasi, dan pembagian menjadi data latih serta data uji. Proses pembersihan melibatkan penghapusan elemen-elemen yang tidak relevan, seperti tanda baca, angka, dan spasi berlebih, untuk meningkatkan kualitas data. Setelah itu, proses tokenisasi mengubah teks menjadi kumpulan kata yang dapat dianalisis oleh model. Data ini kemudian dibagi menjadi dua bagian, yaitu data latih yang digunakan untuk melatih model dan data uji yang digunakan untuk mengukur kinerjanya. Hasil analisis menunjukkan bahwa NAÏVE BAYES mencapai akurasi sebesar 0.6 dengan persentase 60%, performa yang konsisten dalam mengklasifikasikan berbagai kelas sentimen, terutama untuk sentimen netral dan positif. Di sisi lain, KNN menunjukkan akurasi yang jauh lebih rendah, sebesar 0.3 dengan persentase 23%, precision antar kelas yang kurang stabil dimana precision kelas 1 (netral) sangat rendah sebesar 0.21. Perbandingan ini mengindikasikan bahwa NAÏVE BAYES lebih akurat daripada KNN dalam mengidentifikasi sentimen publik terkait proyek IKN, sehingga lebih direkomendasikan untuk digunakan dalam analisis sentimen untuk studi-studi serupa.

Kata kunci: Analisis Sentimen, IKN, K-Nearest Neighbor, NAÏVE BAYES

### ABSTRACT

*The development of Indonesia's new capital city (IKN) has garnered widespread public attention and become a topic of discussion on social media platforms, including YouTube. To understand public sentiment regarding this project, this study employs two machine learning methods, namely K-Nearest Neighbors (KNN) and NAÏVE BAYES. Data were collected from YouTube comments and processed through several stages, including data cleaning (preprocessing), tokenization, and splitting into training and testing datasets. The cleaning process involved removing irrelevant elements such as punctuation marks, numbers, and excessive whitespace to improve data quality. Following this, tokenization transformed the text into a set of words that could be analyzed by the models. The data were then divided into two parts: the training dataset used to train the models and the testing dataset used to evaluate their performance. The analysis results showed that NAÏVE BAYES achieved an accuracy of 0.6 or 60%, demonstrating consistent performance in classifying various sentiment classes, especially for neutral and positive sentiments. On the other hand, KNN exhibited a significantly lower accuracy of 0.3 or 23%, with unstable precision across classes, particularly a very low precision of 0.21 for the neutral class. This comparison indicates that NAÏVE BAYES is more accurate than KNN in identifying public sentiment regarding the IKN project and is therefore recommended for use in similar sentiment analysis studies.*

*Keywords: Sentiment Analysis, National capital, K-Nearest Neighbor, NAÏVE BAYES*

## 1. PENDAHULUAN

### Latar Belakang

Pindahnya Ibu Kota Negara dari DKI Jakarta ke Kalimantan Timur, telah menjadi topik diskusi yang hangat di masyarakat, dengan respons yang beragam mulai dari positif hingga negatif. Pindahan IKN Indonesia memunculkan beragam tanggapan dari masyarakat, mengingat isu ini sangat sensitif dan banyak dibicarakan di media sosial [1]. Analisis sentimen publik terhadap kebijakan ini menjadi penting untuk memahami pandangan masyarakat secara lebih mendalam. Analisis sentimen terhadap opini tentang Ibukota Nusantara penting untuk memahami persepsi masyarakat terkait peran strategisnya sebagai pusat pemerintahan, ekonomi, dan budaya [2].

Penelitian ini memfokuskan pada perbandingan akurasi dua metode klasifikasi, Metode K-Nearest Neighbors (KNN) dan NAÏVE BAYES juga digunakan dalam analisis sentimen publik terhadap pemindahan IKN.

Beberapa penelitian menunjukkan bahwa KNN unggul dalam hal akurasi dan recall dibandingkan Naïve Bayes. Penelitian yang menggunakan KNN untuk menganalisis sentimen terkait pemindahan IKN menunjukkan hasil yang sangat positif, dengan tingkat akurasi 83,75% [3] dan 88,12% [4]. KNN juga mencatatkan presisi 93,98% dan recall 81,53%, menjadikannya pilihan yang lebih efektif dalam menganalisis sentimen terkait IKN dibandingkan Naïve Bayes dalam beberapa konteks ini. Selain itu, KNN memiliki keunggulan dalam menghadapi dataset yang lebih besar dan bervariasi, seperti yang terlihat dalam penelitian yang melibatkan 4.614 data komentar YouTube, di mana KNN menghasilkan akurasi yang lebih tinggi dibandingkan NAÏVE BAYES.

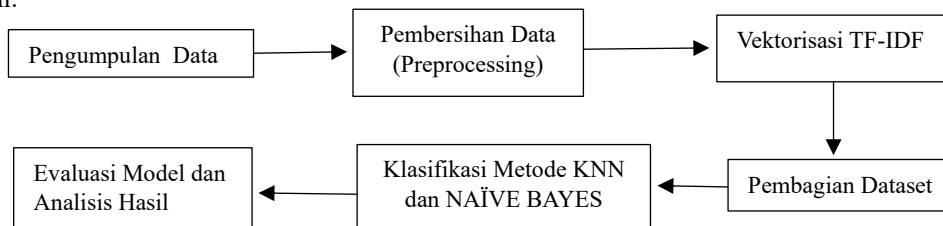
Metode NAÏVE BAYES menunjukkan kinerja bervariasi tergantung pada ukuran dan distribusi data yang digunakan. Dalam beberapa penelitian, seperti yang menganalisis 2.182 tweet, NAÏVE BAYES menunjukkan akurasi 65,26% [5], dan pada dataset yang lebih besar dengan 4.614 data komentar YouTube, akurasi NAÏVE BAYES mencapai 70% [6]. Meskipun kinerjanya lebih rendah dibandingkan dengan KNN dalam hal akurasi, NAÏVE BAYES tetap efektif dalam beberapa kasus, terutama dataset yang lebih kecil. Bahkan, pada penelitian yang menggunakan 546 data, NAÏVE BAYES berhasil memperoleh akurasi 100% pada validasi split dan 90,48% pada cross-validation [7].

Namun, dalam beberapa penelitian lainnya, seperti yang melibatkan perbandingan antara NAÏVE BAYES dan KNN, hasil menunjukkan bahwa KNN lebih unggul dengan akurasi 88,12%, presisi 93,98%, dan recall 81,53%, sementara NAÏVE BAYES mencapai 82,27% untuk akurasi, presisi 86,36%, dan recall 76,93% [4], [8]. Adapun Analisis sentimen terhadap pemindahan ibu kota Indonesia pada Twitter menunjukkan sentimen negatif mendominasi, dengan akurasi tertinggi mencapai 60,606% menggunakan metode Naïve Bayes [9]. Secara keseluruhan, meskipun NAÏVE BAYES efektif untuk analisis sentimen dengan dataset yang lebih kecil dan lebih sederhana, KNN menawarkan performa yang lebih baik dalam konteks analisis sentimen terkait IKN, terutama ketika data lebih besar dan lebih bervariasi.

Berdasarkan analisis berbagai penelitian yang menggunakan KNN dan NAÏVE BAYES untuk klasifikasi sentimen pemindahan IKN, KNN terbukti lebih efektif dalam hal akurasi, presisi, dan recall, terutama ketika jumlah data lebih besar dan kompleks. Sementara itu, NAÏVE BAYES tetap merupakan pilihan yang baik untuk dataset yang lebih kecil atau sederhana, karena kesederhanaannya dalam implementasi. Kemudian pemilihan NAÏVE BAYES juga di perkuat dari penelitian sebelumnya karena kemampuannya yang efektif dalam mengklasifikasikan sentimen, terutama pada dataset yang lebih kecil dan sederhana [10]. Oleh karena itu, pemilihan metode yang tepat sebaiknya mempertimbangkan ukuran dan kompleksitas dataset. Kombinasi kedua metode ini dapat memberikan gambaran yang lebih komprehensif dalam menganalisis persepsi publik terhadap pemindahan IKN, dengan memanfaatkan kekuatan masing-masing algoritma dalam menangani pola kemiripan yang kuat dan distribusi sentimen yang beragam.

## 2. METODE PENELITIAN

Langkah-langkah penelitian ini mengikuti beberapa tahap untuk menganalisis sentimen publik terhadap pembangunan IKN menggunakan metode K-Nearest Neighbors (KNN) dan NAÏVE BAYES. Berikut adalah tahapan yang dilakukan:



Gambar 1. Tahapan Penelitian

### Pengumpulan Data

Pengumpulan data dengan teknik crawling dilakukan melalui API YouTube Data untuk mengakses dan mengunduh komentar dari video tertentu. Setelah berhasil mengumpulkan data, analisis sentimen diterapkan menggunakan VADER Sentiment, yang memberikan skor sentimen untuk mengkategorikan komentar menjadi positif, negatif, atau netral. Selain itu, untuk meningkatkan pemahaman, komentar dapat diterjemahkan ke dalam bahasa yang diinginkan menggunakan Deep Translator sebelum dilakukan analisis sentimen. Pendekatan ini memungkinkan pemahaman yang lebih baik terhadap persepsi publik terhadap konten video.

### Pembersihan Data (*Preprocessing*)

Tahap preprocessing dalam analisis teks mencakup beberapa langkah penting yang bertujuan untuk menyiapkan data agar bisa analisis lebih lanjut. Proses dimulai dengan *tokenisasi*, yang memecah teks menjadi unit-unit kecil yang disebut token, seperti kata atau frasa. Selanjutnya, dilakukan penghapusan *stop words*, yang menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis, sehingga fokus analisis dapat lebih terarah pada kata-kata yang relevan. Kemudian, tahap *stemming* dilakukan untuk mengubah kata menjadi bentuk dasarnya atau akar kata, yang menyatukan variasi kata yang memiliki makna sama. Setelah itu proses normalisasi dimana teks diubah ke dalam format yang konsisten, dengan mengubah huruf besar menjadi huruf kecil atau menghapus tanda baca. Terakhir, pembersihan teks dilakukan untuk menghapus karakter khusus, angka, atau elemen lain yang tidak diperlukan dari teks. Langkah-langkah ini bertujuan untuk peningkatan kualitas dan akurasi analisis teks.

### Vektorisasi TF-IDF

TF-IDF menggabungkan dua komponen: Term Frequency (TF), yang menghitung seberapa sering sebuah kata muncul dalam dokumen tertentu dengan rumus pada Persamaan 1, Inverse Document Frequency (IDF), yang menilai seberapa umum kata tersebut dalam seluruh koleksi dokumen dengan menggunakan rumus pada Persamaan 2. Dengan menggunakan TF-IDF, pengaruh kata-kata umum yang sering muncul dapat dikurangi, sementara kata-kata yang lebih jarang namun signifikan dapat diberi bobot lebih tinggi, membantu model dalam membedakan sentimen dengan lebih efektif menggunakan rumus pada Persamaan 3. Dalam proyek ini, TF-IDF diterapkan pada komentar yang sudah dibersihkan dengan menghitung dua hal utama untuk setiap kata dalam teks: frekuensi kata dalam satu dokumen (TF) dan frekuensi kata dalam seluruh kumpulan dokumen (IDF). Dengan demikian, TF-IDF berperan penting dalam membangun model analisis sentimen yang lebih akurat dengan memprioritaskan kata-kata yang memiliki dampak lebih besar pada hasil analisis.

$$TF(t,d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{jumlah total kata dalam dokumen } d} \quad (1)$$

Keterangan:

- TF mengukur seberapa sering suatu kata (t) muncul dalam dokumen tertentu (d) dibandingkan dengan jumlah total kata dalam dokumen tersebut.
- Memberikan bobot lebih tinggi untuk kata yang sering muncul dalam dokumen, karena kata tersebut cenderung lebih penting untuk memahami isi dokumen.

$$IDF(t,d) = \left( \frac{N}{|\{d \in D : t \in d\}|} \right) \quad (2)$$

Keterangan:

- N adalah jumlah total dokumen dalam korpus atau koleksi dokumen D.
- $|\{d \in D : t \in d\}|$  adalah jumlah dokumen yang mengandung kata t dalam korpus D.

$$TF-IDF(t,d,D) = TF(t,d) \times IDF(t,D) \quad (3)$$

Keterangan:

- TF-IDF adalah hasil perkalian nilai TF dan IDF.

### Pembagian Dataset

Pembagian dataset menjadi data pelatihan dan data pengujian, dimana sebanyak 80% dari data digunakan sebagai data pelatihan, sementara 20% sisanya digunakan sebagai data pengujian. Proporsi ini memungkinkan model untuk belajar dari sebagian besar data yang tersedia dan diuji dengan data yang tidak pernah dilihat sebelumnya agar memberikan hasil uji yang dengan cakupan lebih luas.

### Klasifikasi K-Nearest Neighbor dan Naïve Bayes

NAÏVE BAYES adalah algoritma klasifikasi yang didasarkan pada Teorema BAYES dengan asumsi independensi antar fitur. Algoritma ini menganggap bahwa kehadiran satu fitur tidak bergantung pada fitur lainnya, sehingga setiap fitur berkontribusi mandiri terhadap probabilitas suatu kategori. Misalnya, sebuah buah dapat diklasifikasikan sebagai apel jika memiliki karakteristik tertentu seperti warna merah, bentuk bulat, dan diameter sekitar 3 inci dengan asumsi bahwa setiap karakteristik tersebut independen. Teorema BAYES memberikan cara menghitung probabilitas posterior  $P(c|x)$  dari  $P(c)$ ,  $P(x)$  dan  $P(x|c)$  seperti pada persamaan 4 ini.

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} \quad (4)$$

Keterangan:

- $P(c|x)$  adalah probabilitas posterior kelas ( $c$ , target) yang diberikan prediktor ( $x$ , atribut).
- $P(c)$  adalah probabilitas kelas sebelumnya.
- $P(x|c)$  adalah kemungkinan yang merupakan probabilitas prediktor yang diberikan kelas.
- $P(x)$  adalah probabilitas prediktor sebelumnya.

Dalam analisis sentimen, persamaan ini disesuaikan menjadi persamaan 5.

$$P(\text{sentimen}|\text{teks}) = \frac{P(\text{teks}|\text{sentimen}).P(\text{sentimen})}{P(\text{teks})} \quad (5)$$

K-Nearest Neighbor (KNN) adalah algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi berdasarkan kedekatan atau kemiripan antar data. KNN bekerja dengan prinsip bahwa data yang serupa akan cenderung berada di dekat satu sama lain. Dalam klasifikasi, KNN menentukan kelas dari suatu data baru dengan melihat kelas mayoritas dari sejumlah  $k$  data terdekat (tetangga) di ruang fitur seperti Persamaan 6. KNN dalam analisis sentimen digunakan untuk mengklasifikasikan teks berdasarkan kedekatan atau kemiripan dengan data lain yang sudah diberi label sentimen. Dalam konteks ini, KNN mencari sejumlah tetangga terdekat dari sebuah teks baru dalam ruang fitur dan menentukan sentimen mayoritas dari tetangga tersebut. KNN cocok untuk analisis sentimen karena mampu mengidentifikasi pola-pola sentimen secara intuitif, terutama pada dataset dengan distribusi yang jelas antara sentimen positif, negatif, atau netral.

$$[(x, y), (a, b)] = \sqrt{(x - a)^2 + (y - b)^2} \quad (6)$$

Keterangan:

- $(x,y)$  dan  $(a,b)$  adalah koordinat dari titik.
- $\sqrt{(x - a)^2 + (y - b)^2}$  adalah jarak dihitung dengan mengukur perbedaan posisi pada sumbu  $x$  dan  $y$ , kemudian mengkuadratkan perbedaan tersebut, menjumlahkannya, dan mengambil akar dari hasilnya.

### Evaluasi Model dan Analisis Hasil

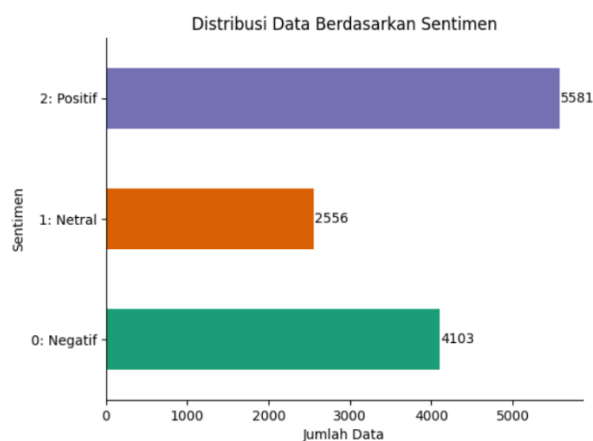
Evaluasi model dilakukan menggunakan metrik seperti akurasi, precision, recall, dan f1-score untuk menilai kinerja model secara menyeluruh. **Akurasi** mengukur persentase prediksi yang benar dari total data. **Precision** mengindikasikan sejauh mana prediksi positif benar-benar relevan, sedangkan **recall** menunjukkan kemampuan model mendeteksi semua sampel yang relevan. **F1-score** adalah rata-rata harmonis antara precision dan recall, menyeimbangkan kedua metrik tersebut terutama saat ada ketidakseimbangan data. Melalui evaluasi ini, hasil kinerja dari berbagai model dibandingkan untuk menentukan metode terbaik yang sesuai dengan karakteristik data dan tujuan analisis. Proses ini membantu mengidentifikasi keunggulan serta kelemahan setiap model, sehingga mendukung pemilihan model yang optimal untuk menghasilkan prediksi sentimen publik yang lebih akurat dan relevan.

## 3. HASIL DAN PEMBAHASAN

### Pengumpulan Data

Data yang didapat dari Youtube dengan total data 12.523, data diolah menggunakan *deep translator* yang digunakan untuk menerjemahkan dari Bahasa Indonesia ke Bahasa Inggris seperti “Yang penting Jokowi kaya” ke “The important thing is that Jokowi is rich”, alasannya karena Pustaka Bahasa Inggris lebih lengkap untuk mengelola sentimen untuk diolah data komentar juga di bersihkan dengan menghilangkan emoji seperti pada komentar “🌈”. Kemudian dianalisis oleh VADER untuk mendapatkan sentimen dari data komen seperti “CANCEL THE IKN, IT IS USELESS. IT WAS BUILT RANDOMLY PLANNING AND IS OF NO USEFUL TO THE PEOPLE.” *Sentiment*

yang diperoleh adalah *negative* dengan threshold 0.05. Ditampilkan banyak data yang berhasil di olah pada gambar 7. yaitu jumlah data per sentiment antara lain; Negative(0) : 4.103; Neutral(1) : 2.556 dan Positive(2) : 5.581 Total Data per Sentimen : 12.240.



Gambar 2. Model Visualisasi Data

### Preprocessing

Proses pembersihan teks dimulai dengan menghapus karakter-karakter khusus yang tidak relevan yang mungkin muncul akibat kesalahan encoding. Selanjutnya, teks dalam kolom yang relevan dibersihkan dengan menghapus tanda baca yang tidak diperlukan, seperti titik dan koma, untuk menjaga keakuratan data. Selanjutnya, teks dikonversi ke huruf kecil guna memastikan konsistensi dan mencegah perbedaan yang tidak diinginkan akibat kapitalisasi. Setelah itu, dilakukan penghapusan kata-kata umum atau stopwords yang tidak memberikan nilai tambah dalam analisis, agar analisis lebih fokus pada kata-kata penting. Proses berikutnya adalah stemming, yaitu mengubah kata menjadi bentuk dasarnya untuk mengurangi variasi kata yang sebenarnya memiliki makna yang sama.

- Mencakup penghapusan karakter khusus “ðŸ”,
- Membersihkan teks pada kolom Comment di dataframe df dengan penghapusan tanda baca seperti tanda “.”
- Pengubahan huruf menjadi huruf kecil serti “Tonton Buka Mata” menjadi “tonton buka mata”,
- Penghapusan kata umum (*stopwords*) seperti “Kami rakyat” menjadi “rakyat”,
- Stemming (pengubahan kata menjadi bentuk dasar) seperti “PENIPUAN” menjadi “tipu”,

### Pemisahan Data

Data dibagi menjadi dua komponen utama: fitur (x), terdiri dari komentar yang telah dibersihkan yaitu “betuulllll moga aja ikn jadi”, dan target (y), merupakan label sentimen (positif, negatif, atau netral) yaitu label “Positive”.

### Vektorisasi TF-IDF

Terdapat sejumlah kata yang sering muncul di banyak dokumen, yang diindikasikan oleh nilai IDF (inverse document frequency) yang rendah. Misalnya, kata-kata seperti “yg” (2.36), “ikn” (2.81), dan “jadi” (2.87) memiliki IDF rendah. Sementara itu, analisis Term Frequency (TF) pada dokumen pertama menampilkan kata-kata dengan frekuensi kemunculan tinggi dalam dokumen tersebut, misalnya “klik” (0.46), “episode” (0.43), dan “lain” (0.30).

### Pembagian Data

Dataset dibagi menjadi dua bagian: set pelatihan (training set) dan set pengujian (testing set) dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Set pelatihan terdiri dari 9.801 contoh data, sementara set pengujian terdiri dari 2.451 contoh.

### Klasifikasi KNN dan Naïve Bayes

Hasil analisis sentimen yang diperoleh menggunakan metode KNN menunjukkan tingkat akurasi yang cukup rendah, yaitu sekitar 23%. Dalam laporan klasifikasi, kinerja model sangat bervariasi antar kelas yaitu:

- Kelas 0 (negative), model memiliki precision sebesar 72%, namun recall-nya hanya 6%, yang berarti meskipun model dapat mengidentifikasi sebagian besar komentar yang seharusnya berlabel 0, hanya sedikit di antaranya yang benar-benar teridentifikasi.

- b) Kelas 1 (netral) menunjukkan kinerja yang lebih baik dengan recall 97%, tetapi memiliki precision yang rendah sebesar 21%, mengindikasikan banyak komentar yang salah dikategorikan sebagai kelas 1.
- c) Kelas 2 (positive) juga menunjukkan precision tinggi sebesar 71%, namun recall hanya 3%, menandakan kesulitan model dalam mendeteksi komentar yang benar-benar berlabel 2.

---

Accuracy K-Nearest Neighbors = 0,23092615259077928

---

Tabel 1. Hasil akurasi metode KNN

Class	Precision	Recall	F1-score	Support
0	0,72	0,06	0,12	844
1	0,21	0,97	0,34	498
2	0,71	0,03	0,05	1108
Accuracy			0,23	2451
Macro Average		0,41	0,26	2451
Weighted Average		0,61	0,23	2451

Hasil analisis sentimen menggunakan metode NAÏVE BAYES menunjukkan peningkatan yang signifikan dalam akurasi dibandingkan dengan metode KNN, dengan tingkat akurasi mencapai sekitar 59.77%. Laporan klasifikasi mengungkapkan variasi kinerja model di antara kelas-kelas yang berbeda antara lain;

- a) Kelas 0 (negative) memiliki precision sebesar 72% dan recall 47%, menandakan bahwa model cukup baik dalam mengidentifikasi komentar yang seharusnya berlabel 0, meskipun masih ada banyak kesalahan identifikasi.
- b) Kelas 1 (netral) menunjukkan hasil yang kurang memuaskan, dengan precision 66% tetapi recall yang sangat rendah, yaitu 12%, yang berarti model kesulitan mendeteksi komentar yang benar-benar berlabel 1.
- c) Kelas 2 (positive) menampilkan kinerja terbaik, dengan recall 91% dan f1-score 69%, menunjukkan bahwa model efektif dalam mengidentifikasi komentar positif di kelas ini.

---

Accuracy NAÏVE BAYES = 0,5977152182782538

---

#### Classification Report

Tabel 2. Hasil akurasi metode Naïve Bayes

Class	Precision	Recall	F1-score	Support
0	0,72	0,47	0,57	844
1	0,66	0,12	0,20	498
2	0,56	0,91	0,69	1108
Accuracy			0,60	2451
Macro Average		0,49	0,37	2451
Weighted Average		0,63	0,60	2451

#### Evaluasi Model dan Analisis Hasil

Berdasarkan analisis sentimen tentang IKN yang dilakukan menggunakan dua metode, yaitu KNN dan NAÏVE BAYES, diperoleh hasil yang menunjukkan perbedaan signifikan dalam akurasi dan kinerja masing-masing metode. KNN menghasilkan akurasi rendah sebesar 23%, dengan precision yang baik untuk beberapa kelas, tetapi kurang efektif dalam mengenali kelas lainnya, khususnya kelas 1. Sementara itu, Naive Bayes menunjukkan akurasi yang jauh lebih baik, yaitu 60%, dengan kemampuan yang lebih konsisten dalam mengklasifikasikan sentimen di berbagai kelas. Precision dan recall yang tinggi untuk kelas tertentu, terutama kelas 2 dengan recall 0.91, mengindikasikan bahwa metode ini lebih efektif dalam menangani dataset yang ada. Secara keseluruhan, hasil analisis ini menegaskan bahwa NAÏVE BAYES lebih unggul dalam memberikan pemahaman yang lebih baik tentang sentimen publik terhadap ibu kota negara dibandingkan dengan KNN.

#### 4. SIMPULAN

Distribusi Sentimen Publik menunjukkan sentimen publik yang beragam terhadap pembangunan IKN, dengan mayoritas sentimen positif atau netral, meski ada kekhawatiran di sebagian kalangan yang perlu diperhatikan

pemerintah. Analisis sentimen dilakukan dengan mengambil data dari youtube menggunakan API dan diklasifikasikan dengan metode KNN dan NAÏVE BAYES. Performa metode berdasarkan analisis dimana NAÏVE BAYES lebih unggul dengan selisih akurasi 37% dalam memetakan sentimen publik terhadap proyek IKN dibandingkan KNN. Ini menegaskan bahwa pendekatan probabilistik seperti Naive Bayes lebih sesuai untuk analisis sentimen pada penelitian ini, serta memberikan pemahaman yang lebih dalam terhadap pandangan masyarakat terkait pembangunan IKN. Selanjutnya, Penelitian mendatang dapat fokus pada pengembangan teknik normalisasi teks untuk memastikan data masukan lebih bersih dan terstruktur, berikutnya memperluas cakupan sumber data, tidak hanya pada YouTube tetapi juga mencakup platform media sosial lainnya seperti Twitter, Instagram, atau forum diskusi publik untuk memperoleh pandangan masyarakat yang lebih luas.

#### DAFTAR PUSTAKA

- [1] S. Lestari, M. Mupaat, dan A. Erfina, “Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter,” *JUSIFO J. Sist. Inf.*, vol. 8, no. 1, hlm. 13–22, Jun 2022, doi: 10.19109/jusifo.v8i1.12116.
- [2] A. Supian, B. Tri Revaldo, N. Marhadi, L. Efrizoni, dan R. Rahmaddeni, “Perbandingan Kinerja Naïve Bayes dan SVM pada Analisis Sentimen Twitter Ibukota Nusantara,” *J. Ilm. Inform.*, vol. 12, no. 01, hlm. 15–21, Mar 2024, doi: 10.33884/jif.v12i01.8721.
- [3] C. Huda dan M. Betty Yel, “Analisa Sentimen Tentang Ibu Kota Nusantara (IKN) dengan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes,” *J. Ilmu Komput. Dan Sist. Inf. JIKOMSI*, vol. 7, no. 1, hlm. 126–130, Feb 2024, doi: 10.55338/jikomsi.v7i1.2846.
- [4] S. D. Prasetyo, S. S. Hilabi, dan F. Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *J. KomtekInfo*, hlm. 1–7, Jan 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [5] J. Muliawan dan E. Dazki, “Sentiment Analysis of Indonesia’s Capital City Relocation Using Three Algorithms: Naïve Bayes, KNN, and Random Forest,” *J. Tek. Inform. Jutif*, vol. 4, no. 5, hlm. 1227–1236, Okt 2023, doi: 10.52436/1.jutif.2023.4.5.1436.
- [6] S. Husyen dan A. Yusnita, “Analisis Sentimen Pemindahan Ibu Kota Negara pada Platform Youtube Menggunakan Metode Naive Bayes Classifier,” vol. 28, no. 2, hlm. 1410–3737, Des 2024, doi: 10.46984/sebatik.v28i2.0000.
- [7] R. K. Septiani, S. Anggraeni, dan S. D. Saraswati, “Klasifikasi Sentimen Terhadap Ibu Kota Nusantara (IKN) pada Media Sosial Menggunakan Naive Bayes,” *TEKNIKA*, vol. 16, no. 02, hlm. 245-254, 2022.
- [8] H. D. A. Assyam dan F. N. Hasan, “Analisis Sentimen Twitter Terhadap Perpindahan Ibu Kota Negara Ke IKN Nusantara Menggunakan Orange Data Mining”. *KLIK*, vol. 4, no. 1, hlm. 341–349, Agu 2023, doi: 10.30865/klik.v4i1.957.
- [9] A. Kurniawan dan S. Waluyo, “Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota pada Twitter,” *Pros. Semin. Nas. Mhs. Fak. Teknol. Inf.*, vol. 1, no. 1, hlm. 455–461, Sep 2022.
- [10] J. Dermawan, Y. Yusra, M. Fikry, S. Agustian, dan L. Oktavia, “Klasifikasi Sentimen Terhadap Topik Pindah Ibu Kota Negara pada Twitter Menggunakan Metode Naïve Bayes Classifier,” *J. Sist. Komput. dan Inform. JSON*, vol. 5, no. 3, hlm. 600, Mar 2024, doi: 10.30865/json.v5i3.7475.