

Sentiment Analysis on the Relocation of the National Capital (IKN) on Social Media X Using Naïve Bayes and K-Nearest Neighbor (KNN) Methods

Nova Wulandari¹, Yana Cahyana², Rahmat³, Hanny Hikmayanti Handayani⁴

Teknik Informatika, Universitas Buana Perjuangan Karawang

[If21.novawulandari@mhs.ubpkarawang.ac.id](mailto:if21.novawulandari@mhs.ubpkarawang.ac.id)¹, yana.cahyana@ubpkarawang.ac.id², rahmat@ubpkarawang.ac.id³,
hanny.hikmayanti@ubpkarawang.ac.id⁴

Article Info

Article history:

Received 2025-04-29

Revised 2025-06-03

Accepted 2025-06-04

Keyword:

Ibu Kota Negara (IKN),
K-Nearest Neighbor,
Naïve Bayes,
Sentiment Analysis,
Social Media X.

ABSTRACT

This study investigates public sentiment toward the relocation of Indonesia's capital from Jakarta to East Kalimantan, focusing on reactions from social media platforms such as X (formerly Twitter). Understanding these sentiments is crucial for the government to gauge support for this significant policy shift. The study compares the performance of two classification algorithms, Naïve Bayes and K-Nearest Neighbor (K-NN), in sentiment analysis. A total of 1.277 comments were collected using the tweet-harvest library through a crawling process. The data underwent preprocessing, including cleaning, case folding, normalization, stopword removal, tokenization, and stemming. Sentiment labels were assigned through both manual and automated methods, while feature extraction was performed using the TF-IDF technique. The algorithms' performance was assessed using accuracy, precision, recall, and F1-score metrics. The results revealed that Naïve Bayes outperformed K-NN, with an accuracy of 70%, precision of 72%, recall of 70%, and an F1-score of 69%. In contrast, K-NN achieved an accuracy of 60%, precision of 62%, recall of 60%, and an F1-score of 59%. These results suggest that Naïve Bayes is more effective in classifying sentiment related to the capital relocation. The findings offer valuable insights for policymakers and highlight the potential of automated sentiment analysis as a tool for monitoring public opinion on major governmental policies.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Pemindahan Ibu Kota Negara (IKN) dari Jakarta ke Kalimantan Timur menjadi strategi pemerintah dalam menghadapi tantangan di Jakarta, termasuk permasalahan lalu lintas yang padat, banjir, dan tingginya kepadatan penduduk. Keputusan ini memicu berbagai tanggapan dari masyarakat, baik yang mendukung maupun menolak, yang banyak diungkapkan melalui media sosial X. Pendapat masyarakat ini memiliki dampak signifikan bagi pemerintah, sehingga diperlukan analisis yang akurat untuk memahami opini publik. Pemindahan Ibu Kota Negara Indonesia mulai diperbincangkan oleh Pemerintah Indonesia pada tahun 2019. Selanjutnya, melalui pertemuan terbatas yang diadakan oleh

Pemerintah Indonesia melalui keputusan Presiden memutuskan untuk pemindahan IKN ke wilayah di luar Pulau Jawa, yang tertera dalam Perencanaan Pembangunan untuk Periode Menengah Nasional untuk Tahun Anggaran 2020-2024 [1].

Presiden Joko Widodo mengumumkan sesi pertemuan pada 29 April 2019 yang membahas tentang relokasi ibu kota negara dari Pulau Jawa. Setelah melalui berbagai penelitian dan dialog masyarakat, akhirnya diputuskan bahwasanya lokasi Ibu kota yang baru akan terletak di Kabupaten Penajam Paser Utara, Kalimantan Timur.[2]. Penentuan lokasi ibu kota baru tentunya didasari oleh pertimbangan dan kajian yang mendalam. Kalimantan Timur dipilih karena dianggap memiliki kemungkinan rendah terhadap bencana alam di

antaranya tanah longsor, banjir, aktivitas vulkanik, gempa, kebakaran hutan maupun tsunami. Selain itu, kawasan ini terletak di pusat Indonesia serra dekat dengan berbagai kota yang sedang tumbuh, seperti Balikpapan maupun Samarinda [3].

Kebijakan ini memicu berbagai tanggapan dari masyarakat, baik yang mendukung maupun menolak, yang banyak diungkapkan melalui media sosial X (sebelumnya Twitter). Media sosial menjadi wadah utama bagi masyarakat dalam menyampaikan opini terkait pemindahan IKN, sehingga analisis sentimen diperlukan untuk memahami persepsi publik. Analisis sentimen merupakan pendekatan yang umum digunakan untuk memahami opini publik terhadap berbagai isu, termasuk kebijakan pemerintah [4]. Namun, penelitian yang secara khusus mengkaji respons publik terhadap pemindahan Ibu Kota Negara (IKN) masih sangat terbatas. Studi ini bertujuan untuk mengisi celah tersebut dengan menganalisis sentimen masyarakat terhadap kebijakan IKN melalui media sosial X (sebelumnya Twitter). Keunggulan penelitian ini terletak pada kombinasi metode pelabelan manual oleh pakar bahasa yang belum banyak diterapkan dalam studi sejenis. Dengan pendekatan ini, penelitian diharapkan mampu memberikan gambaran lebih komprehensif mengenai persepsi publik terhadap kebijakan strategis nasional.

Berbagai penelitian sebelumnya menunjukkan bahwasanya algoritma *Naïve Bayes* maupun KNN telah berhasil diterapkan dalam analisis sentimen pada beragam topik, meskipun dengan tingkat akurasi yang bervariasi. Misalnya, analisis sentimen terhadap komentar pengguna pada video bertema mobil listrik di platform YouTube mengadopsi metode *Naïve Bayes* guna memperoleh tingkat akurasi 70,69%, dengan nilai presisi 43,64% dan *recall* 39,48% [4]. Di sisi lain, penelitian mengenai persepsi publik terhadap dampak politik identitas dalam pemilihan umum 2024 terhadap toleransi antarumat beragama, yang menggunakan pendekatan KNN, memperoleh akurasi sebesar 65% pada subset data tertentu [5]. Penelitian lain yang menganalisis sentimen masyarakat terhadap layanan dari Dinas Kependudukan dan Pencatatan Sipil melalui media sosial Twitter juga menerapkan algoritma *Naïve Bayes* dan menghasilkan akurasi sebesar 65% [6]. Selain itu, kajian tentang tanggapan warganet terhadap kendaraan listrik di platform X dengan pendekatan pembelajaran mesin menunjukkan bahwasanya *Naïve Bayes* memberikan akurasi sebesar 52,9%, dengan presisi 52,1% dan *recall* 49,2%, sedangkan metode KNN menghasilkan akurasi sebesar 49% serta presisi dan *recall*nya yakni masing-masingnya 47,9% [7]. Pada penelitian lain yang membandingkan performa algoritma klasifikasi dalam menganalisis sentimen terhadap isu Islam Nusantara di Indonesia, baik KNN maupun *Naïve Bayes* sama-sama memperoleh akurasi tertinggi sebesar 56,71% [8].

Kajian mengenai analisis sentimen berbasis media sosial telah berkembang pesat dalam beberapa tahun terakhir, terutama dengan memanfaatkan algoritma klasifikasi seperti

Naïve Bayes dan *K-Nearest Neighbor* (KNN). Namun, sebagian besar penelitian tersebut berfokus pada topik umum seperti ulasan produk, opini politik, atau layanan publik, sementara kajian yang secara spesifik mengevaluasi persepsi publik terhadap pemindahan Ibu Kota Negara (IKN) Indonesia masih sangat terbatas. Selain itu, penelitian sebelumnya cenderung mengandalkan data lama dan jarang menggunakan data terbaru yang mencerminkan opini masyarakat saat ini. Perbandingan performa algoritma dalam konteks isu strategis nasional seperti IKN juga belum banyak dieksplorasi secara mendalam, terutama dalam kombinasi dengan pelabelan manual oleh ahli bahasa dan teknik *preprocessing* yang disesuaikan dengan karakteristik bahasa Indonesia. Topik pemindahan Ibu Kota Negara (IKN) dipilih dalam penelitian ini karena merupakan isu strategis nasional yang tengah menjadi perhatian publik luas dan menimbulkan beragam respons masyarakat, khususnya di media sosial. Pemindahan IKN tidak hanya menyangkut aspek geografis, tetapi juga berdampak pada aspek sosial, ekonomi, politik, dan lingkungan. Maraknya diskusi publik mengenai pro dan kontra kebijakan ini menjadikan isu ini penting untuk dianalisis lebih lanjut. Melalui pendekatan analisis sentimen, penelitian ini diharapkan dapat memberikan gambaran yang objektif mengenai tanggapan masyarakat serta memberikan kontribusi dalam pengambilan kebijakan yang lebih responsif dan berbasis data.

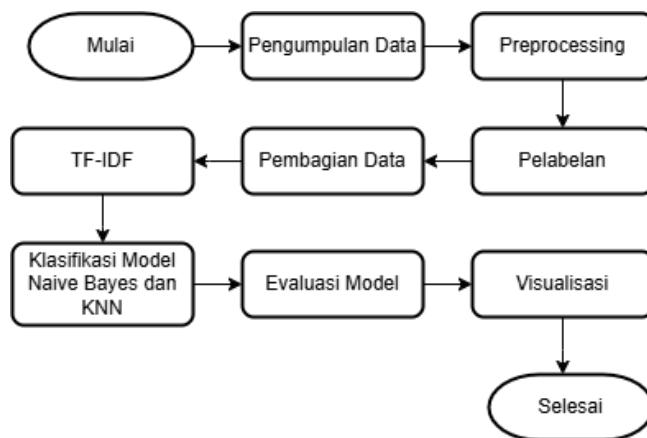
Adapun tujuannya pada penelitian ini guna menganalisis sentimen masyarakat terkait kebijakan pemindahan Ibu Kota Negara (IKN) serta membandingkan kinerja algoritma *Naïve Bayes* maupun *K-Nearest Neighbor* (KNN) dengan menggunakan metrik evaluasi di antaranya yaitu akurasi, presisi, recall, maupun F1-score. Diharapkan hasil dari penelitian ini dapat memberikan pemahaman mengenai respons publik dan mendukung pengembangan metode analisis sentimen di masa yang akan datang.

II. METODE

Dalam penelitian ini, algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dipilih karena keduanya merupakan metode klasifikasi yang umum digunakan dalam analisis sentimen dan memiliki karakteristik berbeda yang saling melengkapi. *Naïve Bayes*, yang berbasis probabilistik, sangat cocok untuk data teks berdimensi tinggi seperti hasil TF-IDF karena efisien dalam proses pelatihan dan tetap memberikan akurasi yang baik meskipun asumsi independensi antar fitur tidak sepenuhnya terpenuhi. Di sisi lain, KNN digunakan sebagai pembanding karena mengklasifikasikan data berdasarkan kemiripan dengan tetangga terdekat, meskipun performanya cenderung menurun pada data berdimensi tinggi akibat *curse of dimensionality*. Meskipun begitu, pendekatan KNN tetap relevan karena tidak memerlukan proses pelatihan eksplisit dan mencerminkan distribusi data aktual [5]. Implementasi kedua algoritma ini memungkinkan evaluasi efektivitas pendekatan probabilistik dan berbasis kemiripan dalam

menganalisis opini publik terhadap pemindahan Ibu Kota Negara (IKN).

Langkah-langkah dalam penelitian ini mengikuti diagram alur yang menggambarkan setiap tahap yang dilakukan selama proses penelitian. Berikut adalah gambar alur penelitian.



Gambar 1. Alur Penelitian

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari media sosial Twitter (X) melalui proses *crawling* menggunakan pustaka *tweet-harvest* dengan fitur *auto token*. Pengumpulan data dilakukan dalam dua tahap, yaitu pada tanggal 15 Maret 2025 hingga 18 Maret 2025 yang menghasilkan sebanyak 810 komentar, dan pada tanggal 5 Mei 2025 hingga 7 Mei 2025 yang menghasilkan 467 komentar. Secara keseluruhan, jumlah data yang berhasil dihimpun adalah 1.277 komentar. Data tersebut disimpan dalam format CSV dan memuat informasi seperti ID tweet, *username*, waktu unggahan, isi komentar, jumlah *like*, jumlah *retweet*, serta *hashtag* yang digunakan.

Pemilihan akun pengguna Twitter dilakukan secara acak tanpa adanya kriteria khusus, sehingga data yang diperoleh mencerminkan keragaman opini dari berbagai pengguna. Pendekatan ini dimaksudkan untuk menghindari kencenderungan terhadap akun tertentu dan memastikan bahwa analisis sentimen dilakukan terhadap sampel data yang lebih objektif dan representatif.

B. Preprocessing

Dalam penelitian ini, tahap pra-pemrosesan dilakukan dengan beberapa langkah, yaitu: menghapus URL, angka, tanda baca, emoji, dan spasi berlebih; melakukan *case folding*; normalisasi kata tidak baku; *stopword removal*; *tokenisasi*; serta *stemming*[6]. Seluruh tahapan ini bertujuan untuk mempersiapkan data agar lebih bersih dan konsisten sebelum dilakukan analisis sentimen.

C. Pelabelan

Proses pelabelan data dilakukan dengan cara manual oleh ahli bahasa yaitu Adelya Daniyah, M.Pd yang memperhatikan

konteks kalimat serta arti kata secara keseluruhan. Setiap teks diperiksa secara teliti untuk mengelompokkan sentimennya ke dalam tiga kelas, yakni positif, negatif, ataupun netral. Metode ini diterapkan untuk memastikan keakuratan label yang digunakan dalam proses pelatihan model.

D. Pembagian Data

Dalam penelitian ini, data terbagi atas dua bagian yakni latih maupun uji dengan perbandingannya 80% berbanding 20%. Setelah melalui tahap *preprocessing*, jumlah data yang tersedia adalah 1.112. Dari jumlah tersebut, sebanyak 889 data digunakan sebagai data latih, Sementara itu, 223 data digunakan sebagai data uji, Pembagian ini dirancang untuk melatih model secara efisien dan menilai performa algoritma *Naïve Bayes* serta KNN dalam melakukan analisis sentimen yang berhubungan dengan pemindahan Ibu Kota Negara (IKN).

E. Pembobotan TF-IDF

TF-IDF sebagai teknik yang digunakan untuk mengevaluasi signifikansi suatu kata dalam sebuah dokumen, dengan memperhatikan frekuensi kemunculannya di dalam dokumen tersebut dan frekuensi kemunculan kata tersebut di berbagai dokumen lainnya [7].

F. Pelatihan model Naïve Bayes dan KNN

Pelatihan model dilakukan menggunakan *Naïve Bayes* dan KNN untuk mengklasifikasikan sentimen terkait pemindahan IKN. *Naïve Bayes* Bekerja dengan menghitung probabilitas kata untuk setiap kelas sentimen (positif, negatif, netral) menggunakan *Teorema Bayes*, kemudian memprediksi sentimen dengan memilih kelas yang memiliki probabilitas tertinggi [8]. Berikut adalah rumus *Naïve Bayes*.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan:

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis bahwa data X merupakan suatu *class* spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (*posteriori probability*)
- $P(H)$: Probabilitas hipotesis H (*prior probability*)
- $P(X|H)$: Probabilitas X berdasarkan kondisi hipotesis H
- $P(X)$: Probabilitas dari X

Sementara itu, KNN menentukan sentimen berdasarkan mayoritas K tetangga terdekat menggunakan *Euclidean Distance*, di mana model mencari kesamaan teks berdasarkan jarak antar data. Performa K-NN bergantung pada pemilihan nilai K yang optimal [9].

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \quad (2)$$

Keterangan:

d : Jarak
 a : Data uji/testing
 b : Sample Data
 I : Variable Data
 n : Dimensi Data

Kedua model dievaluasi menggunakan 223 data uji dan dibandingkan berdasarkan metrik akurasi, presisi, *recall*, serta *F1-score* guna mengetahui metode yang paling optimal dalam melakukan analisis sentimen terhadap IKN.

G. Evaluasi Model

Confusion matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja algoritma klasifikasi dengan membandingkan hasil prediksi model terhadap label yang sebenarnya. Matriks ini sangat berguna dalam mengidentifikasi sejauh mana model mampu mengklasifikasikan data secara akurat, terutama dalam masalah klasifikasi biner. Dalam *confusion matrix* terdapat empat komponen utama, yaitu True Positive (TP), yang menunjukkan jumlah data yang benar-benar positif dan diprediksi positif; True Negative (TN), yaitu data yang benar-benar negatif dan diprediksi negatif; False Positive (FP), yaitu data yang sebenarnya negatif namun diprediksi positif; serta False Negative (FN), yaitu data yang sebenarnya positif tetapi diprediksi negatif. Melalui keempat nilai ini, dapat dihitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1-score*.

Pada penelitian ini, proses *confusion matrix* bertujuan untuk evaluasi model *Naïve Bayes* dan KNN dalam analisis sentimen pemindahan ibu kota negara di platform X. Berikut rumus *confusion matrix*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = 2 \times \frac{precision \times recall}{precision + recall} \quad (6)$$

H. Visualisasi

Selanjutnya, dilakukan analisis visual menggunakan *Word Cloud*, metode visualisasi ini menggambarkan berbagai kata yang paling sering muncul dalam sebuah teks, di mana kata dengan frekuensi kemunculan yang lebih tinggi akan diperbesar ukurannya, sehingga mempermudah identifikasi tema atau kata kunci utama dalam data [11].

III. HASIL DAN PEMBAHASAN

A. Crawling Data

Penelitian ini menggunakan dataset dari *Twitter* untuk analisis sentimen terkait IKN. Proses pengambilan data dilakukan dengan *Google Colab* melalui bahasa pemrograman *Python*. Pada *crawling* data dari *Twitter*, digunakan auto token yang memungkinkan interaksi dengan platform tanpa perlu registrasi manual. Total data yang berhasil dikumpulkan sebanyak 1.277 komentar, yang kemudian ditampilkan pada Gambar 2.

conversation_id_str	created_at	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	lang
0 1901956167586476134	Tue Mar 18 11:17:41 +0000 2025	0	Negara menggunakan pola represif pada negara negara	1901956167586476134	NaN	NaN	i
1 1901955344856952871	Tue Mar 18 11:11:25 +0000 2025	0	yuk kita sebarkan kebenaran biar IKN bebas dari berita palsu https://t.co/X5PjIzO9Ug	1901955344856952871	https://pbs.twimg.com/media/GmUaTerI0iPjzP.jpg	NaN	i
2 1901955046163824867	Tue Mar 18 11:13:14 +0000 2025	1	Kapan pindah ke IKN buruan deh pindah sono siapa tahu bisa mengurangi kemacetan Jabodetabek https://t.co/6EqouT5Ei1	1901955046163824867	https://pbs.twimg.com/media/GmUaBtEAAK_L0.jpg	NaN	i

Gambar 2. Hasil *Crawling* dataset

B. Preprocessing

1. Cleaning Data

Tahap pertama, *Cleaning*, mencakup penghapusan URL, angka, tanda baca, emoji, spasi berlebihan, serta karakter *non-Alphanumeric*. Proses ini dilakukan untuk memastikan bahwasanya analisis hanya berfokus pada teks yang memiliki makna [12]. Berikut adalah hasil *Cleaning* data.

TABEL I
HASIL CLEANING DATA

Sebelum	Sesudah
yuk kita sebarkan kebenaran biar IKN bebas dari berita palsu https://t.co/X5PjIzO9Ug	yuk kita sebarkan kebenaran biar IKN bebas dari berita palsu
Kapan pindah ke IKN buruan deh pindah sono siapa tahu bisa mengurangi kemacetan Jabodetabek https://t.co/6EqouT5Ei1	Kapan pindah ke IKN buruan deh pindah sono siapa tahu bisa mengurangi kemacetan Jabodetabek

2. Case Folding

Case folding merupakan suatu proses dalam pengolahan teks yang mana tujuannya guna menyamakan format penulisan dengan mengubah keseluruhan huruf dalam teks menjadi huruf kecil (*lowercase*) [13].

TABEL II
HASIL CASE FOLDING

Sebelum	Sesudah
yuk kita sebarkan kebenaran biar IKN bebas dari berita palsu	yuk kita sebarkan kebenaran biar ikn bebas dari berita palsu
Kapan pindah ke IKN buruan deh pindah sono siapa tahu bisa mengurangi kemacetan jabodetabek	kapan pindah ke ikn buruan deh pindah sono siapa tahu bisa mengurangi kemacetan jabodetabek

3. Normalisasi

Tahap normalisasi dilakukan untuk memastikan setiap kata sesuai dengan ejaan dalam KBBI serta mengoreksi kata-kata tidak baku menjadi bentuk yang benar [14].

TABEL III
HASIL NORMALISASI

Sebelum	Sesudah
yuk kita sebarkan kebenaran biar ikn bebas dari berita palsu	yuk kita sebarkan kebenaran biar ikn bebas dari berita palsu
kapan pindah ke ikn buruan deh pindah sono siapa tahu bisa mengurangi kemacetan jabodetabek	kapan pindah ke ikn buruan deh pindah sana siapa tahu bisa mengurangi kemacetan jabodetabek

4. Stopword

Stopword dilakukan untuk menghilangkan kata-kata yang umum dan tidak bermakna [15].

TABEL IV
HASIL STOPWORD

Sebelum	Sesudah
yuk kita sebarkan kebenaran biar ikn bebas dari berita palsu	kita sebarkan kebenaran biar ikn bebas berita palsu
kapan pindah ke ikn buruan deh pindah sana siapa tahu bisa mengurangi kemacetan jabodetabek	kapan pindah ikn buruan pindah sana siapa tahu mengurangi kemacetan jabodetabek

5. Tokenisasi

Tokenisasi adalah proses memecah teks atau kalimat menjadi unit-unit terkecil, seperti kata, frasa, atau simbol yang disebut token.

TABEL V
HASIL TOKENISASI

Sebelum	Sesudah
kita sebarkan kebenaran biar ikn bebas berita palsu	kita, sebarkan, kebenaran, biar, ikn, bebas, berita, palsu
kapan pindah ikn buruan pindah sana siapa tahu mengurangi kemacetan jabodetabek	kapan, pindah, ikn, buruan, pindah, sana, siapa, tahu, mengurangi, kemacetan, jabodetabek

6. Stemming

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya dengan cara menghapus imbuhan [16].

TABEL VI
HASIL STEMMING

Sebelum	Sesudah
kita, sebarkan, kebenaran, biar, ikn, bebas, berita, palsu	sebar benar biar ikn bebas berita palsu
kapan, pindah, ikn, buruan, pindah, sana, siapa, tahu, mengurangi, kemacetan, jabodetabek	kapan pindah ikn buru pindah sana siapa tahu kurang macet jabodetabek

C. Labeling

Proses pelabelan data dilakukan dengan cara manual oleh ahli bahasa yaitu Adelya Daniyah, M.Pd yang memperhatikan konteks kalimat serta arti kata secara keseluruhan. Setiap teks diperiksa secara teliti untuk mengelompokkan sentimennya ke dalam tiga kelas, yakni positif, negatif, ataupun netral. Metode ini diterapkan untuk memastikan keakuratan label yang digunakan dalam proses pelatihan model. Adapun hasil pelabelan yang dilakukan ahli bahasa disajikan berikut ini.

TABEL VII
DATA HASIL PAKAR

No	TEXT	LABEL
1	lapor malaikat maut biar skalian cabut nyawa mulai jabat lengser kok bikin ramai aja mulai soal esemka bangun ikntidak pakai uang apbn deret tipu.	Negative
2	tetangga w milih suami biar pindah ikn bang.	Positive
3	ikn simbol harap baru perata ekonomi indonesia.	Positive

Berikut hasil sentimen positif, negatif dan netral.

TABEL VIII
HASIL SENTIMEN

Sentimen	Jumlah Tweet
Positif	474
Negatif	339
Netral	299
Total	1.112

Dari total 1.112 tweet yang dianalisis, sebanyak 474 (42,6%) tergolong positif, 339 (30,5%) negatif, dan 299 (26,9%) netral. Data ini menunjukkan adanya ketidakseimbangan sentimen, di mana tweet positif mendominasi. Kondisi ini perlu diperhatikan karena dapat memengaruhi akurasi analisis, terutama jika digunakan untuk pelatihan model sentimen otomatis.

D. Pembagian Data

Dataset dibagi menjadi dua kelompok utama, yakni 80% untuk data pelatihan maupun 20% bagi data pengujian. Sebanyak 892 sampel digunakan untuk melatih model, sementara 223 sampel sisanya digunakan untuk mengevaluasi kinerja model tersebut. Pembagian ini bertujuan untuk memberikan jumlah data yang cukup dalam proses pelatihan

serta untuk mengevaluasi kemampuan model dalam mengklasifikasikan data yang belum pernah dikenali sebelumnya.

E. Pembobotan TF-IDF

Pembobotan TF-IDF dalam penelitian ini dilakukan guna mengidentifikasi kata-kata yang memiliki pengaruh besar dalam dokumen. Hasil tertinggi menunjukkan kata "IKN" dengan skor 18.76, "bangun" 15.80, dan "nusantara" 14.40. Kata-kata tersebut memiliki frekuensi yang tinggi dan bobot yang signifikan dalam konteks data yang dianalisis. Berikut hasil TF-IDF.

	Kata	TF-IDF	Frekuensi
663	i kn	18.762209	116
1107	nusantara	15.809387	102
854	kota	14.406261	110
655	ibu	13.008668	96
1531	the	10.966644	145

Gambar 3. Hasil TF-IDF

F. Klasifikasi Model Algoritma Naïve Bayes dan KNN

Pada tahap ini, model *machine learning* dibangun dengan menggunakan data latih dan uji yang dipilih secara acak dari keseluruhan dataset. Langkah ini bertujuan untuk melakukan validasi silang (*cross validation*) serta memperoleh nilai akurasi prediksi. Penelitian ini mengimplementasikan algoritma *Naïve Bayes* dan KNN. Pada tahap awal klasifikasi, dataset dibagi menjadi 80% bagi latih serta 20% untuk uji, dengan tujuan untuk mengevaluasi kinerja kedua algoritma tersebut.

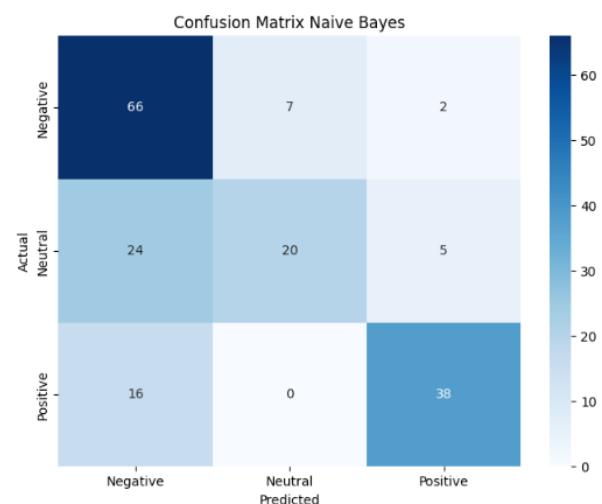
Klasifikasi menggunakan *Naïve Bayes* memperoleh *accuracy* sebesar 70%, *precision* 72%, *recall* 70%, *f1-score* 69%. Proses ini mencakup *splitting* data latih dan uji, ekstraksi fitur, pelatihan model, evaluasi model, dan pembuatan *confusion matrix*. Berikut detail hasil evaluasi dengan menerapkan model *Naïve Bayes*.

Evaluasi Model Naïve Bayes:				
Classification Report:				
	precision	recall	f1-score	support
Negative	0.62	0.88	0.73	75
Netral	0.74	0.41	0.53	49
Positive	0.84	0.70	0.77	54
accuracy			0.70	178
macro avg	0.74	0.66	0.67	178
weighted avg	0.72	0.70	0.69	178

Accuracy Score: 0.6966292134831461

Gambar 4. Hasil Klasifikasi *Naïve Bayes*

Pada *confusion matrix* untuk model *Naïve Bayes*, hasil klasifikasi menunjukkan performanya yang cukup baik dengan akurasinya sebesar 70%, presisi 72%, recall 70%, maupun *F1-score* 69%. Nilai-nilai ini mencerminkan kemampuan model dalam mengklasifikasikan data sentimen secara cukup konsisten. Adapun berikut ini adalah tampilan *confusion matrix* yang dihasilkan oleh model *Naïve Bayes*:



Gambar 5. *Confusion Matrix Naïve Bayes*

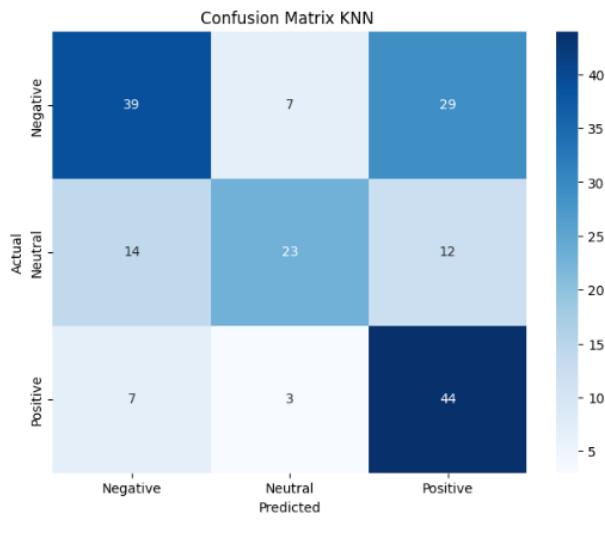
Sementara itu, klasifikasi menggunakan KNN menghasilkan akurasi sebesar 60%, presisi 62%, *recall* 60%, dan *F1-score* 59%. Proses yang dilakukan serupa, meliputi pembagian data latih dan uji, ekstraksi fitur, pelatihan model, evaluasi model, dan pembuatan *confusion matrix*. Berikut adalah rincian hasil evaluasi dengan model KNN.

KNN - Classification Report:				
	precision	recall	f1-score	support
Negative	0.65	0.52	0.58	75
Netral	0.70	0.47	0.56	49
Positive	0.52	0.81	0.63	54
accuracy			0.60	178
macro avg	0.62	0.60	0.59	178
weighted avg	0.62	0.60	0.59	178

KNN - Accuracy Score:
0.5955056179775281

Gambar 6. Hasil Klasifikasi KNN

Pada *confusion matrix* untuk model *K-Nearest Neighbor* (KNN), hasil klasifikasi menunjukkan bahwasanya model memiliki akurasi sebesar 60%, presisi 62%, *recall* 60%, maupun *F1-score* 59%. Nilai ini mengindikasikan bahwasanya performa model KNN dalam mengklasifikasikan data sentimen masih kurang optimal jika dibandingkan dengan model *Naïve Bayes*. Berikut adalah tampilan *confusion matrix* yang diperoleh dari hasil pengujian model KNN:



Gambar 7. Confusion matrix KNN

Berikut adalah hasil perbandingan model *Naïve Bayes* dan KNN.

TABEL IX

HASIL PERBANDINGAN MODEL ALGORITMA NAÏVE BAYES DAN KNN

Model	Akurasi	Presisi	Recall	F1-Score
<i>Naïve Bayes</i>	70%	72%	70%	69%
KNN	60%	62%	60%	59%

Berdasarkan hasil yang ditunjukkan pada Tabel 9, algoritma *Naïve Bayes* memperlihatkan performa yang lebih baik dengan akurasi 70%, presisi 72%, *recall* 70%, dan nilai *F1-score* 69%. Sementara itu, algoritma KNN mempunyai kinerja yang lebih rendah dengan akurasi 60%, presisi 62%, *recall* 60%, dan *F1-score* sebesar 59%. Berdasarkan perbandingan tersebut, dapat disimpulkannya bahwasanya *Naïve Bayes* lebih unggul dibandingkan *K-Nearest Neighbor* (KNN) dalam analisis sentimen terkait pemindahan Ibu Kota Negara (IKN). *Naïve Bayes* efektif untuk data teks berdimensi tinggi, seperti hasil ekstraksi fitur TF-IDF, karena menggunakan pendekatan probabilistik dengan asumsi independensi antar fitur. Hal ini membuatnya lebih efisien dalam mengelola data teks dan memberikan akurasi yang tinggi. Sementara itu, KNN mengalami kesulitan dengan banyaknya fitur karena bergantung pada perhitungan jarak antar data, yang terpengaruh oleh *curse of dimensionality*. *Naïve Bayes* juga lebih cepat karena pelatihan dilakukan sekali, dan prediksi hanya melibatkan perhitungan probabilitas, sedangkan KNN memerlukan pencarian seluruh data latih. Selain itu, *Naïve Bayes* lebih tahan terhadap noise dan data tidak relevan, sedangkan KNN sangat bergantung pada data tetangga terdekat yang dapat kurang representatif

G. Visualisasi

Dalam penelitian ini, *Wordcloud* berperan sebagai alat visualisasi teks yang menampilkan kata-kata dengan variasi ukuran dan warna, berdasarkan seberapa sering dan seberapa penting kata tersebut muncul dalam konteks analisis. Alat ini

digunakan untuk menggambarkan tweet positif, negatif dan netral yang dominan [17]. Berikut hasil visualisasi.



Gambar 8. Hasil Wordcloud

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan pemindahan Ibu Kota Negara (IKN) melalui media sosial X menggunakan dua algoritma klasifikasi: *Naïve Bayes* dan *K-Nearest Neighbor* (KNN). Berdasarkan 1.112 data komentar yang telah diberi label secara manual, ditemukan bahwa sentimen positif mendominasi (42,6%), diikuti oleh sentimen negatif (30,5%) dan netral (26,9%). Hasil ini menunjukkan bahwa meskipun banyak dukungan terhadap relokasi IKN, terdapat pula kekhawatiran signifikan yang perlu diperhatikan oleh pembuat kebijakan.

Dalam proses klasifikasi sentimen, algoritma *Naïve Bayes* menunjukkan performa yang lebih unggul dibandingkan KNN. *Naïve Bayes* berhasil mencapai akurasi sebesar 70%, sedangkan KNN hanya mencapai 60%. Perbedaan ini disebabkan oleh keunggulan *Naïve Bayes* dalam menangani data teks berdimensi tinggi dan kemampuannya memberikan prediksi yang lebih stabil melalui pendekatan probabilistik. Di sisi lain, KNN memiliki keterbatasan dalam konteks data besar karena bergantung pada perhitungan jarak yang rawan terpengaruh oleh jumlah fitur dan noise pada data. Secara umum, opini publik yang positif mencerminkan harapan terhadap pemerataan pembangunan dan pemerintahan yang lebih efisien di luar Pulau Jawa. Namun, sentimen negatif menggarisbawahi kekhawatiran mengenai pembiayaan proyek, dampak sosial, dan risiko lingkungan. Sementara itu, komentar netral mengindikasikan adanya kelompok masyarakat yang masih bersikap hati-hati dalam menyikapi kebijakan ini.

Dari hasil penelitian ini, dapat disimpulkan bahwa metode *Naïve Bayes* lebih efektif untuk diterapkan dalam analisis sentimen terkait isu strategis nasional seperti pemindahan IKN. Temuan ini diharapkan dapat memberikan kontribusi bagi pemerintah dalam memahami opini publik secara lebih akurat dan mendalam, serta menjadi dasar untuk pengambilan keputusan dan perumusan kebijakan yang lebih responsif terhadap suara masyarakat.

DAFTAR PUSTAKA

[1] S. Lestari *et al.*, “Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter,” vol. 8, no. 1, pp. 13–22, doi: <https://doi.org/10.19109/jusifo.v8i1.12116>.

[2] C. Huda and M. Betty Yel, “Analisa Sentimen Tentang Ibu Kota Nusantara (IKN) Dengan Menggunakan Algoritma K-Nearest Neighbors (KNN) dan Naïve Bayes,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI) V*, vol. 7, no. 1, pp. 126–130, 2024, doi: <https://doi.org/10.55338/jikomsi.v7i1.2846>.

[3] S. I. R. Adi, B. Bakkara, K. A. Zega, F. N. Vielita, and N. A. Rakhmawati, “Analisis Sentimen Masyarakat Terhadap Progress Ikn Menggunakan Model Decision Tree,” *JIK4 (Jurnal Informatika)*, vol. 8, no. 1, p. 57, Jan. 2024, doi: 10.31000/jika.v8i1.9803.

[4] A. M. Pravina, K. N. Sani, H. D. Harfianto, T. A. Pratama, A. Fahrina, and Y. Ruldeviyani, “Sentiment Analysis Of Delivery Service Opinions On Twitter Documents Using K-Nearest Neighbor Arsyia, et, al [Sentiment Analysis Of Delivery Service Opinions On Twitter Documents Using K-Nearest Neighbor],” *Hal; Jl. Salemba Raya*, vol. 9, no. 2, pp. 996–1012, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>

[5] H. Wisnu, M. Afif, and Y. Ruldevyani, “Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naïve Bayes,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1742-6596/1444/1/012034.

[6] M. Fikri Haikal and J. Indra, “Analisis Sentimen Bakal Calon Presiden Indonesia 2024 Dengan Algoritma Naïve Bayes”.

[7] Y. Khoiruddin, A. Fauzi, and A. M. Siregar, “Analisis Sentimen Gojek Indonesia Pada Twitter Menggunakan Algoritme Naïve Bayes Dan Support Vector Machine”.

[8] R. Saputra and F. N. Hasan, “Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naïve Bayes,” *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 6, no. 3, pp. 411–419, Jul. 2024, doi: 10.47233/jtekstis.v6i3.1378.

[9] A. J. Arifin and A. Nugroho, “Uji Akurasi Penggunaan Metode KNN dalam Analisis Sentimen Kenaikan Harga BBM pada Media Twitter”, doi: <http://dx.doi.org/10.35889/progresif.v19i2.1288>.

[10] T. Cahya Herdiyani and A. U. Zailani, “Sentiment Analysis Terkait Pemindahan Ibu Kota Indonesia Menggunakan Metode Random Forest Berdasarkan Tweet Warga Negara Indonesia Sentiment Analysis Related to Transportation of Indonesian Capital City Using Random Forest Method Based On Tweet Of Indonesian Citizens,” 2022.

[11] Tania Puspa Rahayu Sanjaya, Ahmad Fauzi, and Anis Fitri Nur Masruriyah, “Analisis sentimen ulasan pada e-commerce shopee menggunakan algoritma naive bayes dan support vector machine,” *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 16–26, Jun. 2023, doi: 10.37373/infotech.v4i1.422.

[12] Y. Cahyana and A. M. Siregar, “Analisis Sentiment Pembelajaran Tatap Muka Terbatas (PTMT) Selama Pandemik Covid-19 Menggunakan Algoritma Naïve Bayes.” doi: <https://doi.org/10.33322/petir.v16i2.1964>. “2191-Article Text-6789-1-10-20230205”.

[13] A. Azhar, S. U. Masruroh, L. K. Wardhani, and O. Okfalisa, “Performance comparison of the Naïve Bayes algorithm and the k-NN lexicon approach on Twitter media sentiment analysis,” *Science, Technology and Communication Journal*, vol. 3, no. 2, pp. 33–38, Feb. 2023, doi: 10.59190/stc.v3i2.229.

[14] 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: <https://doi.org/10.37373/tekno.v10i2.419>.

[15] C. Atika Sari and E. Hari Rachmawanto, “Sentiment Analyst on Twitter Using the K-Nearest Neighbors (KNN) Algorithm Against Covid-19 Vaccination” 2022.

[16] M. Padhy, U. M. Modibbo, R. Rautray, S. S. Tripathy, and S. Bebortta, “Application of Machine Learning Techniques to Classify Twitter Sentiments Using Vectorization Techniques,” Aug. 20, 2024. doi: 10.20944/preprints202408.1395.v1.

[17]