

Sentiment Analysis on Public Perception of the Nusantara Capital on Social Media X Using Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) Methods

Dinda Haliza^{1*}, Muhammad Ikhsan^{2*}

* Universitas Islam Negeri Sumatera Utara

dindhilzaa@gmail.com¹, mhd.ikhsan@uinsu.ac.id²

Article Info

Article history:

Received 2025-03-20

Revised 2025-04-08

Accepted 2025-04-14

Keyword:

*Sentiment Analysis,
Confusion Matrix,
IKN,
K-Nearest Neighbor,
Support Vector Machine.*

ABSTRACT

The relocation and development of the National Capital City (IKN) as the center of government activities has become a hot topic, sparking diverse opinions among the public. The proposal to move the capital from DKI Jakarta to East Kalimantan has drawn significant attention from online communities, particularly on social media platform X (Twitter). This study aims to explore public sentiment regarding the development of IKN by applying artificial intelligence-based classification algorithms, namely Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). Sentiments are categorized as positive or negative to provide deeper insights into public perceptions. Through web crawling techniques, a total of 4,000 data points were collected. After the preprocessing stage, 3,608 data points remained, which were then translated into English to facilitate labeling using the Vader Sentiment method. The analysis results indicate that negative sentiment (1,873) is more dominant than positive sentiment (1,735). The data was then split into two sets: 80% for training (2,886 data points) and 20% for testing (722 data points). Based on the evaluation results, SVM and K-NN proved to be effective for sentiment analysis. SVM achieved an accuracy of 76%, precision of 78%, recall of 81%, and an f1-score of 79%, while K-NN attained an accuracy of 65%, precision of 62%, recall of 98%, and an f1-score of 76%. With superior performance, SVM emerges as a more reliable method for classifying public sentiment regarding the IKN development policy.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Pemindahan ibu kota negara Indonesia dari Jakarta ke daerah tertentu di Kalimantan Timur diatur dalam Undang-Undang Nomor 3 Tahun 2022 tentang Ibu Kota Negara (IKN). Setelah disahkan DPR pada 18 Januari 2022, Presiden Joko Widodo menandatangani menjadi undang-undang pada 22 Januari 2022. Ibu Kota Negara Indonesia disebutkan berada di Kecamatan Samboja, Kabupaten Kutai Kartanegara, Kecamatan Sepak, Kabupaten Penajam Paser Utara, Kalimantan Timur [1]. Kalimantan Timur merupakan lokasi yang ideal untuk Ibu Kota Negara Baru karena relatif bebas dari bahaya kenaikan muka air laut, gempa bumi, tsunami, kebakaran hutan, letusan gunung berapi, dan tanah longsor. Kota-kota seperti Balikpapan dan Samarinda yang

berkembang pesat melengkapi posisi Indonesia yang menguntungkan di wilayah tengah [2].

Di era digital saat ini, penerapan kebijakan pemerintah menjadi topik yang sering dibahas di berbagai jejaring sosial, terutama di *Twitter* (X). Media ini memungkinkan penggunaannya untuk berbagi komentar, opini, pandangan, cerita, ide, serta berita terkini tentang berbagai hal yang menarik perhatian publik. Selain itu, *Twitter* (X) juga sering dijadikan sumber data untuk berbagai penelitian sosial, termasuk analisis sentimen [3].

Analisis sentimen merujuk pada proses untuk mengidentifikasi kecenderungan atau perasaan seseorang berdasarkan teks atau tulisan, yang biasanya ditemukan di media sosial. Sentimen ini dapat bernilai positif, negatif, atau netral. Proses ini umumnya dilakukan secara otomatis menggunakan berbagai aplikasi atau media yang

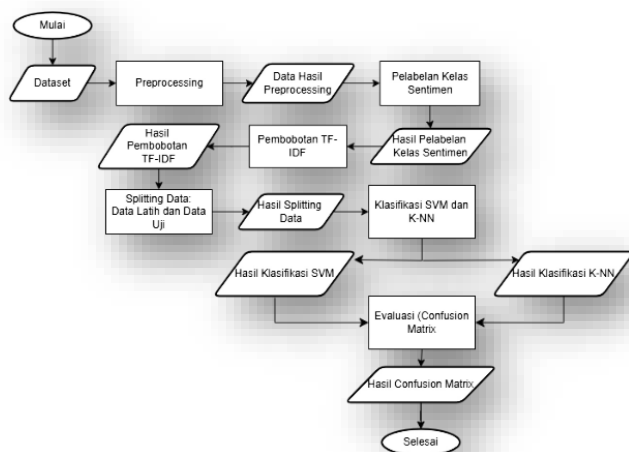
mendukung analisis teks [4]. Tujuan dari analisis sentimen adalah untuk menentukan apakah suatu teks cenderung mengarah ke polaritas positif atau negatif. Klasifikasi teks dari postingan di media sosial selalu menjadi topik penelitian yang menarik dan penuh tantangan [5].

Support Vector Machines (SVM) mempunyai konsep matematis yang lebih kuat dan terdefinisi dengan baik dalam model klasifikasi. SVM digunakan untuk menentukan *hyperplane* optimal dengan memaksimalkan jarak antara kelas-kelas yang ada [6]. Kemudian, K-Nearest Neighbor (K-NN) adalah metode kategorisasi objek yang memperhitungkan tetangga terdekat objek dan frekuensi kemunculannya untuk sampai pada kesimpulan akhir [7].

Penelitian ini bertujuan mengevaluasi akurasi SVM dan K-NN dalam menganalisis data dari media sosial X (Twitter) serta mengidentifikasi sentimen positif dan negatif. Studi ini berfokus pada persepsi masyarakat Indonesia terhadap pemindahan Ibu Kota Nusantara, dengan menyoroti sentimen terkini yang dipengaruhi oleh peristiwa, perubahan kebijakan, atau faktor lain yang membentuk opini publik.

Pemilihan SVM dan K-NN didasarkan pada efisiensi, interpretabilitas, dan kesesuaian dengan dataset. SVM unggul dalam klasifikasi teks dengan margin maksimal antar kelas, sementara K-NN digunakan sebagai pembanding sederhana. *Naïve Bayes* tidak digunakan karena keterbatasan asumsi independensi fitur, dan *Random Forest* lebih cocok untuk data numerik. *Deep Learning* (LSTM, BERT) memang lebih akurat dalam NLP, tetapi membutuhkan dataset besar, komputasi tinggi, dan waktu pelatihan lama. Oleh karena itu, SVM dan K-NN dipilih sebagai metode yang lebih praktis dan efisien untuk analisis sentimen dalam penelitian ini.

II. METODE



Gambar 1 Alur Penelitian

A. Dataset

Penelitian ini menggunakan 4000 data *tweet* yang dikumpulkan dengan menggunakan bahasa Indonesia dan bertaggar *ikn*. Proses pengambilan data dilakukan melalui *library Harvest*, yang memerlukan akses *auth_token* yang dapat diperoleh dari situs *web* pengembang *Twitter*. Dengan menggunakan *auth_token*, *library* ini mendapatkan izin untuk mengakses API yang disediakan oleh *Twitter*. Melalui API tersebut, fitur pencarian pada *library* ini memungkinkan pencarian *tweet* yang relevan dengan kata kunci yang ditentukan.

B. Pre-Processing

Langkah awal yang penting dalam penambangan data adalah pra-pemrosesan, yang melibatkan transformasi data mentah ke dalam format yang dapat digunakan oleh prosedur penambangan [8]. Langkah ini mempersiapkan data untuk penggunaan pada tahap-tahap berikutnya dengan memproses data teks untuk meningkatkan kualitasnya [9]. *cleaning*, *case folding*, *stopword removal*, *stemming*, dan *tokenization* merupakan bagian dari proses pra-pemrosesan ini. [10].

C. Pelabelan Data

Pelabelan data adalah proses mengidentifikasi dan menandai sampel data yang umum digunakan dalam konteks pelatihan model *machine learning*. Proses *labelling* data dilakukan secara otomatis menggunakan metode *vader* yang merupakan bahasa pemrograman *python* dari fitur *NLTK* (*Natural Language Toolkit*) menghasilkan kelas sentimen berdasarkan *lexicon* dengan *source* Inggris [11].

D. TF-IDF

Kata-kata yang telah diproses terlebih dahulu diberi nilai pada tahap berikutnya yang disebut pembobotan kata. Tujuan dari prosedur ini adalah untuk menetapkan kepentingan relatif untuk setiap kata sebelum menggunakannya sebagai masukan klasifikasi. Pendekatan TF-IDF digunakan untuk pembobotan dalam penelitian ini. Dengan melihat gambaran keseluruhan, pendekatan TF-IDF mencoba untuk menentukan seberapa penting sebuah kata dalam sebuah dokumen [12].

$$tf = 0,5 + 0,5 \times \frac{tf}{\max(tf)} \dots\dots\dots (1)$$

$$idf = \log\left(\frac{D}{df_i}\right) \dots\dots\dots (2)$$

$$Wd.t = tf.d.t \times idf.d.t \dots\dots\dots (3)$$

Keterangan :

D = dokumen ke-d

t = term ke-t dari dokumen

W = bobot ke-d terhadap term ke-t

tf = frekuensi kemunculan term i dalam dokumen

idf = *Inversed Document Frequency*

df = jumlah dokumen yang mengandung term i

E. Splitting Data

Membagi dataset menjadi dua akan memudahkan Anda mengelola set pelatihan dan pengujian. Untuk analisis ini, proporsi data pelatihan terhadap data pengujian adalah 80:20. Sebanyak 3608 titik data, termasuk 2886 titik data untuk pelatihan dan 722 titik data untuk pengujian, telah diekstraksi dari 4000 dataset yang telah menjalani praproses.

F. Support Vector Machine (SVM)

Support vector machines (SVM) adalah jenis *machine learning* yang mencari jarak *hyperplane* optimal antara dua kelas. Kelas 1 dan kelas -1 digunakan untuk memisahkan anggota metode SVM [13]. Dalam *support vector machines (SVM)*, pemisahan antara kelas positif dan negatif dimaksimalkan dengan menentukan *hyperplane* terbaik. [14].

G. K-Nearest Neighbors (K-NN)

Salah satu pendekatan untuk klasifikasi objek yang memperhitungkan titik data yang secara geografis paling dekat dengan item adalah *K-Nearest Neighbors (K-NN)*. Dengan menggunakan fitur dan sampel pelatihan, sistem ini dapat mengkategorikan berbagai hal. Untuk memperkirakan kelas dari contoh kueri baru, metode K-NN menggunakan kedekatan titik sebagai basisnya. Algoritma K-NN menggunakan teknik berikut untuk menemukan jarak kedekatan data [15].

$$d = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots \dots \dots (4)$$

Keterangan :

d : jarak
a : data uji
b : sample data
i : variable data
n : dimensi data

H. Confusion Matrix

Performa model klasifikasi dapat dinilai menggunakan matriks konfusi. Perbandingan antara kelas data aktual dan temuan kelas yang diproyeksikan dilakukan dalam matriks konfusi. Sejumlah ukuran, termasuk *recall*, akurasi, presisi, dan *f-score*, dihitung dari perbandingan ini [16]. Mengukur akurasi dan performa model klasifikasi yang diterapkan merupakan tujuan dari metode penilaian ini [17]. Dari persamaan lima hingga delapan, kita dapat melihat rumus matriks konfusi. [18].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \dots \dots \dots (5)$$

$$Recall = \frac{TP}{TP+FN} \dots \dots \dots (6)$$

$$Presicion = \frac{TP}{TP+FP} \dots \dots \dots (7)$$

$$F1-Score = \frac{2*Recall*Presicion}{Recall+preicion} \dots \dots \dots (8)$$

Penjelasan:

TP (*True Positive*): Data positif yang diberi label secara akurat.

TN (*True Negative*): Data negatif yang terklasifikasi dengan akurat.

FP (*False Positive*): Data negatif yang terklasifikasi sebagai akurat.

FN (*False Negative*): Data positif yang terklasifikasi sebagai tidak akurat.

III. HASIL DAN PEMBAHASAN

A. Dataset

Data merupakan hasil *crawling* pada media sosial X dengan kata kunci “ikn” dengan mengumpulkan 4000 data mentah. Pada tahap *crawling* mengumpulkan data dalam bentuk teks yang diekstrak dari tweet yang dikirim oleh pengguna media sosial X, tanpa memandang usia mereka. Kemudian menyimpan data mentah ini dalam format *Comma Separated Values (CSV)*.

TABEL I
KUMPULAN DATASET

No.	Text
1	ga masalah si dimana aja yang penting ga ikn (maaf kalimantan)
2	Video Full Tom Lembong ttg IKN Mitos Pertumbuhan Ekonomi Indonesia 5% Kenaikan Angka Pengangguran UMKM dll
3	@sinopalkan @txttrpemerintah Kayak Di ikn. Kalo org yg kerja disana mikirin duit iya emang menghasilkan. Tapi kalo yg mikirin kesehatan fisik dan mental pada ngeluh. Ini saya tanya 2 org yg kerja disana ya. Tekanannya ngeri. Semua serba ngebut pengerjaannya. Demi upacara 17 agustus tahun
...	...
4000	@liputan6dotcom Pak jokowi emang jenius sekali bisa menekan biayaa demi ikn

B. Preprocessing

Setelah data berhasil dikumpulkan, selanjutnya akan diolah ke tahap *preprocessing* data. *preprocessing* merupakan tahap seleksi pengecekan teks dengan cara membersihkan teks, memperbaiki kesalahan pada teks dan menyederhanakan teks untuk tahap selanjutnya. *Preprocessing* bertujuan untuk memperbaiki kualitas teks pada data. *Preprocessing* meliputi *cleaning*, *case folding*, *normalization*, *tokenizing*, *stopword removal*, *stemming*.

TABEL 2
HASIL PREPROCESSING

Cleaning	ga masalah si dimana aja yang penting ga ikn maaf kalimantan
Case Folding	ga masalah si dimana aja yang penting ga ikn maaf kalimantan
Normalization	tidak masalah sih dimana saja yang penting tidak ikn maaf kalimantan
Tokenizing	['tidak', 'masalah', 'sih', 'dimana', 'saja', 'yang', 'penting', 'tidak', 'ikn', 'maaf', 'kalimantan']
Stopword Removal	['masalah', 'dimana', 'yang', 'penting', 'ikn', 'maaf', 'kalimantan']
Stemming	masalah mana yang penting ikn maaf kalimantan

C. Pelabelan Sentimen

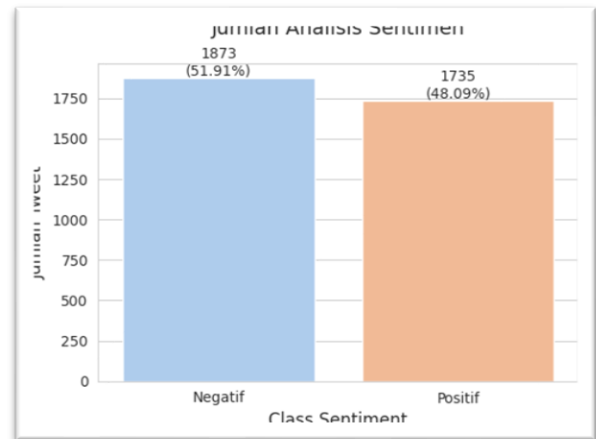
Proses pelabelan dilakukan menggunakan metode VADER (*Valence Aware Dictionary and sEntiment Reasoner*). Untuk meningkatkan akurasi hasil analisis sentimen, data teks yang awalnya dalam Bahasa Indonesia diterjemahkan terlebih dahulu ke dalam Bahasa Inggris. Hal ini dilakukan agar VADER dapat lebih optimal dalam mengkategorikan sentimen berdasarkan *lexicon* bawaan yang dirancang untuk Bahasa Inggris. Setelah diterjemahkan, VADER akan menentukan kelas sentimen (positif atau negatif) berdasarkan nilai skor *lexicon* dan memberikan skor sentimen total yang disebut "*compound score*", yang mencerminkan kecenderungan sentimen dalam teks. Seperti yang ditunjukkan pada Tabel 3.

TABEL 3
HASIL PREPROCESSING

Sentimen	Positive	Negative	Compound	Label
Which problem is important IKN sorry Kalimantan	0.184	0.408	-0.2960	Negatif

TABEL 4
JUMLAH KELAS SENTIMEN

Label	Jumlah
Positif	1.735
Negatif	1.873
Total	3.608



Gambar 2. Hasil Pelabelan Sentimen

Berdasarkan hasil dari pelabelan tersebut terdiri dari 3.608 data sentimen telah dibagi menjadi beberapa kelas sentimen dengan total 1.735 kelas positif dan 1.873 kelas negatif. Oleh karena itu banyak pengguna X yang memberikan respon negatif terkait perkembangan IKN.

Dataset ini tidak sepenuhnya seimbang, meskipun perbedaannya tidak terlalu mencolok. Namun, karena selisih jumlah antara kedua kelas tidak terlalu besar, dampaknya terhadap analisis kemungkinan tidak signifikan. Mayoritas sentimen yang muncul bersifat negatif, yang terutama dipicu oleh kekhawatiran terkait lingkungan, ekonomi, dan sosial. Masyarakat menyoroti isu deforestasi, pembengkakan anggaran negara, ketidakpastian investasi, serta dampak perubahan budaya dan demografi. Untuk mengurangi sentimen negatif ini, pemerintah perlu meningkatkan sosialisasi, menerapkan kebijakan yang ramah lingkungan, memberikan dukungan ekonomi, dan memastikan pembangunan yang merata. Dengan langkah-langkah ini, kepercayaan publik terhadap pemindahan IKN dapat ditingkatkan.

D. TF-IDF

Setelah pelabelan, tahap pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) dilakukan. Di sini, data dokumen dievaluasi secara numerik untuk menetapkan nilai pada setiap kata (istilah). Pembobotan ini bertujuan untuk mengukur seberapa penting suatu kata dalam dokumen dibandingkan dengan seluruh kumpulan dokumen. Model klasifikasi analisis sentimen akan dilatih menggunakan contoh perhitungan nilai TF dan DF yang diberikan untuk 5 data pelatihan dan 2 data uji. Tabel 5 menunjukkan temuan pembobotan TF-IDF, yang digunakan untuk meningkatkan efektivitas representasi teks dalam proses klasifikasi.

TABEL 5
HASIL TF-IDF SVM

TER M	TF-IDF				
	D1	D2	D3	D4	D5
Tolak	1,4771 21	0	0	0	0
Ikn	5	5	5	5	5
Perintah	0	1,4771 21	0	0	0
Terus	0	1,4771 21	0	0	0
Bangunan	0	1,4771 21	0	0	0
Lebih	0	1,4771 21	0	0	0
Maju	0	1,4771 21	0	0	0
Hebat	0	0	1,4771 21	0	0
Harus	0	0	0	1,4771 21	0
Coret	0	0	0	1,4771 21	0
Hutan	0	0	0	0	1,4771 21
Gundul	0	0	0	0	1,4771 21
Buat	0	0	0	0	1,4771 21

Tahap berikutnya adalah normalisasi nilai TF-IDF untuk menyelaraskan skala pada setiap data. Normalisasi dilakukan agar seluruh nilai pada dataset memiliki rentang yang seragam, sehingga analisis menjadi lebih akurat dan terhindar dari bias.

TABEL 6
HASIL NORMALISASI SVM

No	D1	D2	D3	D4	D5
1	0,1201 34	0	0	0	0
2	0,4066 48	0,406648	0,4066 48	0,406648	0,406648
3	0	0,120133 789	0	0	0
4	0	0,120134	0	0	0
5	0	0,120134	0	0	0
6	0	0,120134	0	0	0
7	0	0,120133 789	0	0	0
8	0	0	0,1201 34	0	0
9	0	0	0	0,120133 789	0
10	0	0	0	0,120133 789	0
11	0	0	0	0	0,120134
12	0	0	0	0	0,120133 789

No	D1	D2	D3	D4	D5
13	0	0	0	0	0,120133 789

Setelah mendapatkan hasil TF-IDF untuk klasifikasi SVM, kemudian menghitung Kembali nilai TF-IDF untuk klaifikasi K-NN yang dimana nilai data latih dan data uji dihitung berbarengan.

TABEL 7
HASIL TF-IDF K-NN

TER M	TF-IDF						
	D1	D2	D3	D4	D5	D6	D7
Tolak	1,60 2059	0	0	0	0	0	0
ikn	7	7	7	7	7	7	7
perintah	0	1,60 2059	0	0	0	0	0
terus	0	1,60 2059	0	0	0	0	0
bangun	0	1,60 2059	0	0	0	0	0
lebih	0	1,60 2059	0	0	0	0	0
maju	0	1,60 2059	0	0	0	0	0
hebat	0	0	1,60 2059	0	0	0	0
harus	0	0	0	1,60 2059	0	0	0
coret	0	0	0	1,60 2059	0	0	0
hutan	0	0	0	0	1,60 2059	1,60 2059	0
gundul	0	0	0	0	1,60 2059	0	0
buat	0	0	0	0	1,60 2059	0	0
jadi	0	0	0	0	0	1,60 2059	0
kota	0	0	0	0	0	1,60 2059	0
nusantara	0	0	0	0	0	0	1,60 2059
bawak	0	0	0	0	0	0	1,60 2059
damak	0	0	0	0	0	0	1,60 2059
baik	0	0	0	0	0	0	1,60 2059
indonesia	0	0	0	0	0	0	1,60 2059

Tahap berikutnya adalah normalisasi nilai TF-IDF untuk menyelaraskan skala pada setiap data. Normalisasi dilakukan agar seluruh nilai pada dataset memiliki rentang yang seragam, sehingga analisis menjadi lebih akurat dan terhindar dari bias.

TABEL 8
HASIL NORMALISASI K-NN

NO	D1	D2	D3	D4	D5	D6	D7
1	0,08 068	0	0	0	0	0	0
2	0,35 250	0,35 250	0,35 250	0,352 50	0,352 50	0,352 50	0,352 50
3	0	0,08 068	0	0	0	0	0
4	0	0,08 068	0	0	0	0	0
5	0	0,08 068	0	0	0	0	0
6	0	0,08 068	0	0	0	0	0
7	0	0,08 068	0	0	0	0	0
8	0	0	0,08 068	0	0	0	0
9	0	0	0	0,080 68	0	0	0
10	0	0	0	0,080 68	0	0	0
11	0	0	0	0	0,080 68	0,080 68	0
12	0	0	0	0	0,080 68	0	0
13	0	0	0	0	0,080 68	0	0
14	0	0	0	0	0	0,080 68	0
15	0	0	0	0	0	0,080 68	0
16	0	0	0	0	0	0	0,080 68
17	0	0	0	0	0	0	0,080 68
18	0	0	0	0	0	0	0,080 68
19	0	0	0	0	0	0	0,080 68
20	0	0	0	0	0	0	0,080 68

E. Support Vector Machine (SVM)

Menemukan nilai kernel untuk setiap set pengujian menggunakan set pelatihan adalah langkah pertama dalam memanfaatkan *Support Vector Machine*. Memperoleh nilai kernel selanjutnya menghitung bobot data pengujian. Nilai fungsi $f(x)$ kemudian ditentukan dengan menggunakan bobot yang dihasilkan. Data akan dikategorikan sebagai positif berdasarkan *hyperplane* yang dipilih jika memenuhi persamaan $w.x + b = 1$. Sebaliknya, data akan diklasifikasikan sebagai negatif jika persamaan $w.x + b = -1$ terpenuhi.

1. Data uji 1

["ikn", "jadi", "kota", "hutan"]

$$f(x) = w.x + b$$

$$= \sum \alpha_i y_i K(x_i, x_j) + b$$

$$= (0,1928 + 0,1714 + 0,1714 + 0,1424) + (-0,0061)$$

$$= 0,68398$$

$$\text{Fungsi Klasifikasi} = \text{sign}(0,68398)$$

$$= 1$$

2. Data uji 2

["ikn", "nusantara", "bawa", "dampak", "baik", "indonesia"]

$$f(x) = w.x + b$$

$$= \sum \alpha_i y_i K(x_i, x_j) + b$$

$$= (0,1928 + 0,1714 + 0,1714 + 0,1424 + 0,1424 + 0,1424) + (-0,0061)$$

$$= 1,05574$$

$$\text{Fungsi Klasifikasi} = \text{sign}(1,05574)$$

$$= 1$$

Setelah dilakukan pengujian terhadap dua data uji, hasil menunjukkan bahwa pada fungsi klasifikasi, kedua data uji memiliki nilai $f(x) = 1$. Berdasarkan aturan klasifikasi, nilai ini menunjukkan bahwa kedua data tersebut termasuk dalam kategori kelas 1, yang merepresentasikan kelas Positif.

F. K-Nearest Neighbor

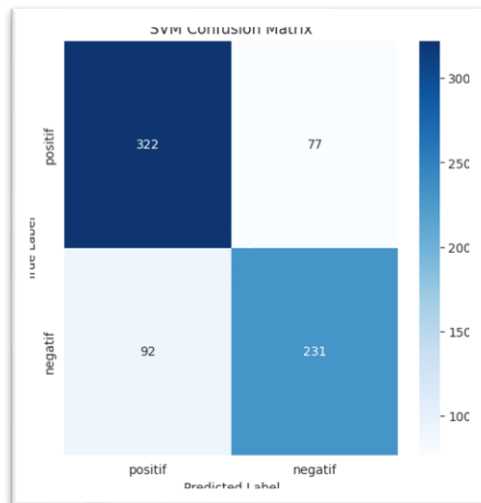
Pertama dalam menggunakan *K-Nearest Neighbor* adalah menentukan nilai K. Dalam penelitian ini, nilai K ditetapkan pada 3 ($K = 3$). Pengukuran jarak antara data pelatihan dan data uji digunakan untuk melakukan komputasi. Selanjutnya, menggunakan $K = 3$, peringkat jarak ditentukan dari yang terkecil hingga terbesar berdasarkan tiga nilai terkecil. Seperti yang ditunjukkan pada Tabel 9, prosedur ini menjamin bahwa pilihan kategorisasi didasarkan pada tiga tetangga terdekat yang paling relevan.

TABEL 9
HASIL KLASIFIKASI K-NN

Data Uji	Label
["ikn", "jadi", "kota", "hutan"]	Negatif
["ikn", "nusantara", "bawa", "dampak", "baik", "indonesia"]	Negatif

G. Confusion Matrix

Evaluasi hasil penelitian ini dilakukan dengan menggunakan *Confusion Matrix*. Melalui *Confusion Matrix*, nilai *accuracy*, *precision*, *recall*, dan *f1-score* akan dihitung untuk membangun model klasifikasi yang menggunakan metode SVM dan K-NN.



Gambar 3. Confusion Matrix SVM

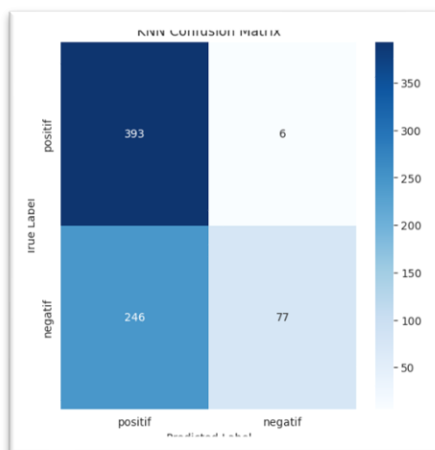
Berdasarkan Gambar 3. Dengan menerapkan rumus berikut pada data uji, kita dapat menentukan hasil uji klasifikasi *Support Vector Machine (SVM)*, termasuk nilai akurasi, presisi, *recall*, dan *f1* skor.

$$Accuracy = \frac{322+231}{322+77+92+231} \times 100\% = 76\%$$

$$Precision = \frac{322}{322+92} \times 100\% = 78\%$$

$$Recall = \frac{322}{322+77} \times 100\% = 81\%$$

$$F1-score = \frac{2 \times 78 \times 81}{78+81} \times 100\% = 79\%$$



Gambar 4. Confusion Matrix K-NN

Berdasarkan Gambar 4. dapat diperoleh hasil terkait pengujian klasifikasi *K-Nearest Neighbor (KNN)* terhadap data uji, nilai *accuracy*, *precision*, *recall*, dan *f1*-score akan dihitung menggunakan persamaan berikut.

$$Accuracy = \frac{393+77}{393+6+246+77} \times 100\% = 65\%$$

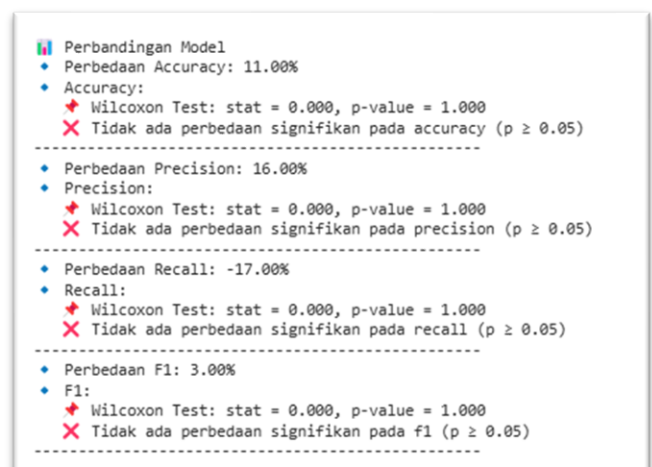
$$Precision = \frac{393}{393+246} \times 100\% = 62\%$$

$$Recall = \frac{393}{393+6} \times 100\% = 98\%$$

$$F1-score = \frac{2 \times 62 \times 98}{62+98} \times 100\% = 76\%$$

K-NN memiliki *recall* tinggi tetapi *precision* rendah karena kecenderungannya mengklasifikasikan data berdasarkan mayoritas K tetangga terdekat. *Recall* tinggi terjadi karena model berhasil mengenali hampir semua sampel positif, sementara *precision* rendah disebabkan oleh banyaknya *False Positives* (FP), di mana data negatif salah diklasifikasikan sebagai positif. Hal ini bisa terjadi jika K terlalu kecil ($K = 3$) atau jika distribusi data tidak seimbang. Untuk mengatasi masalah ini, dapat digunakan nilai K yang lebih besar, metode *weighted voting*, atau penyeimbangan data latih.

SVM lebih seimbang dalam *precision* dan *recall* karena mampu menggeneralisasi data lebih baik, lebih tahan terhadap *outlier*, dan tidak hanya mengandalkan kedekatan lokal seperti K-NN. Dengan *margin* optimal, SVM mengurangi *False Positives* dan *False Negatives* secara bersamaan, sementara K-NN lebih rentan terhadap distribusi data yang tidak seimbang dan *curse of dimensionality*. Hal ini membuat SVM lebih andal dalam tugas *sentiment analysis* dibandingkan K-NN.



Gambar 5 Perbandingan Model

Berdasarkan Gambar 5, hasil perbandingan model menunjukkan selisih sebesar 11%. Pengujian menggunakan *Wilcoxon Test* mengindikasikan bahwa perbedaan tersebut tidak signifikan secara statistik. Dengan demikian, tidak terdapat bukti kuat untuk menyatakan bahwa salah satu model memiliki performa yang secara signifikan lebih baik dibandingkan yang lain.

IV. KESIMPULAN

Dengan menerapkan algoritma *Support Vector Machine* dan *K-Nearest Neighbor* dengan kuantitas data sentimen yang diperoleh selama proses *crawling* sebanyak 4000 data terhadap perkembangan IKN pada media sosial X. 3608 data yang tersisa setelah *processing*. Dengan menggunakan pendekatan *Vader Sentiment* untuk pelabelan, diperoleh bahwa kelas sentimen negatif memiliki keunggulan 1873 dibandingkan dengan kelas sentimen positif 1735. Jelas dari data ini bahwa pengguna sosial X lebih cenderung memberikan umpan balik negatif dalam menilai kemajuan IKN. Data yang digunakan untuk pengujian dalam penelitian ini terdiri dari 2886 titik data latih dan 722 titik data uji, dengan total 80% data latih dan 20% data uji. *f1-score* 79%, *recall* 81%, presisi 78%, dan akurasi 76% semuanya dicapai dalam klasifikasi *Support Vector Machine*. Namun *K-Nearest Neighbor* mencapai akurasi 65%, presisi 62%, *recall* 98%, dan *f1-score* 76%.

DAFTAR PUSTAKA

- [1] Munawaroh, A., Ridhoi, R., & Rudiman, R. (2024). Sentiment Analysis Dengan Naïve Bayes Berbasis Orange Terhadap Resiko Pembangunan IKN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 587-592.
- [2] Adi, S. I. R., Bakkara, B., Zega, K. A., Vielita, F. N., & Rakhmawati, N. A. (2024). Analisis Sentimen Masyarakat Terhadap Progres IKN Menggunakan Model Decision Tree. *JIKA (Jurnal Informatika)*, 8(1), 57-65.
- [3] Puspita, R., & Suciati, T. N. (2020). Mobile phone dan media sosial: Penggunaan dan tantangannya pada jurnalisme online Indonesia. *Ekspres Dan Persepsi: Jurnal Ilmu Komunikasi*, 3(2), 132-146. M. Wegmuller, J. P. von der Weid, P. Oberson, & N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," *Prosiding ECOC'00*, 2000, paper 11.3.4, p. 109.
- [4] Syah, H., & Witanti, A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm). *Jurnal Sistem Informasi Dan Informatika (Simika)*, 5(1), 59-67.
- [5] Lestandy, M., Abdurrahim, A., & Syafa'ah, L. (2021). Analisis Sentimen Tweet Vaksin COVID-19 Menggunakan Recurrent Neural Network dan Naïve Bayes. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(4), 802-808.
- [6] Ardiansyah, D., Saepudin, A., Aryanti, R., & Fitriani, E. (2023). Analisis Sentimen Review Pada Aplikasi Media Sosial Tiktok Menggunakan Algoritma K-Nn Dan Svm Berbasis Pso. *Jurnal Informatika Kaputama (JIK)*, 7(2), 233-241.
- [7] Legito, L., Riau, N. P., Putro, A. N. S., Mardiani, E., Arifin, N. Y., Sepriano, S., & Erkamim, M. (2023). Penerapan Algoritma K-Nearest Neighbor untuk Analisis Sentimen Terhadap Isu Khilafah dan Radikalisme di Indonesia: Implementation K-Nearest Neighbor Algorithm for Sentiment Analysis on Khilafah and Radicalism Issues in Indonesia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 324-330.
- [8] Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan metode naïve bayes dan support vector machine pada analisis sentimen twitter. *SMATIKA Jurnal: STIKI Informatika Jurnal*, 10(02), 71-76.
- [9] Khairunnisa, S., Adiwijaya, A., & Al Faraby, S. (2021). Pengaruh Text Preprocessing terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19). *J. Media Inform. Budidarma*, 5(2), 406.
- [10] Irfham, L. G., Adiwijaya, A., & Wisesty, U. N. (2019). Klasifikasi Berita Bahasa Indonesia Menggunakan Mutual Information dan Support Vector Machine. *J. Media Inform. Budidarma*, 3(4), 284.
- [11] Anwar, M. T. (2023). Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 10(1), 783-792.
- [12] Septiani, D., & Isabela, I. (2022). Analisis term frequency inverse document frequency (tf-idf) dalam temu kembali informasi pada dokumen teks. *SINTESIA J. Sist. dan Teknol. Inf. Indones*, 1(2), 81-88.
- [13] Ramlan, R., Satyahadewi, N., & Andani, W. (2023). Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM. *Jambura Journal of Mathematics*, 5(2), 431-445.
- [14] Fitriyah, N., Warsito, B., & Di Asih, I. M. (2020). Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *Jurnal Gaussian*, 9(3), 376-390.
- [15] Mara, A. T., Sedyono, E., & Purnomo, H. (2021). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Metode Pembelajaran Dalam Jaringan (Daring) Di Universitas Kristen Wira Wacana Sumba. *JOINTER: Journal of Informatics Engineering*, 2(01), 24-31.
- [16] Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18-26.
- [17] Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan akurasi dan waktu proses algoritma K-NN dan SVM dalam analisis sentimen twitter. *Jurnal Informatika*, 6(2), 226-235.
- [18] Bria, N. A. R., & Witanti, A. (2023). Analisis Sentimen Masyarakat Indonesia Menggunakan Algoritma Support Vector Machine Tentang Pilpres 2024. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3328-3333.