

# Sentiment Analysis of Indonesian Responses to the Conflict in Palestine Using KNN and SVM Methods

Rizky Fauzi<sup>1\*</sup>, Erik Iman Heri Ujjianto<sup>2\*</sup>

\*Program Studi Informatika, Universitas Teknologi Yogyakarta, Indonesia

\*\* Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia  
[rizkyfauzii2003@gmail.com](mailto:rizkyfauzii2003@gmail.com)<sup>1</sup>, [erik.iman@uty.ac.id](mailto:erik.iman@uty.ac.id)<sup>2</sup>

## Article Info

### Article history:

Received 2024-10-22

Revised 2024-10-28

Accepted 2024-10-31

### Keyword:

*Palestine,*

*Israel,*

*Sentiment,*

*K-Nearest Neighbors (KNN),*

*Twitter.*

## ABSTRACT

The prolonged conflict between Palestine and Israel has attracted worldwide attention, including Indonesia, which has a history of strong support for the Palestinian cause. This study aims to analyze the sentiment of Indonesian people towards the Palestinian-Israeli conflict using the K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) methods. The subject of this research is user data X (Twitter) which contains opinions about the conflict. After preprocessing, weighting, and labeling, 2960 tweets were collected and classified into three sentiment categories: positive, negative, and neutral. The KNN+SVM method is applied to classify the sentiment of the processed tweet data. The results showed that of the 2960 data analyzed, 33.8% were labeled positive, 38.9% were labeled negative, and 27.4% were labeled neutral with 82% accuracy, 83% precision, 82% recall, and 82% F1-Score. These results show that the majority of Indonesians tend to be negative in expressing their views on the Palestinian-Israeli conflict. This analysis provides greater insight into sentiment patterns in Indonesian responses to sensitive issues, and contributes to the study of public opinion and social dynamics on social media.



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

## I. PENDAHULUAN

Perselisihan yang berlangsung lama antara Palestina dan Israel telah menarik perhatian dunia secara luas, termasuk di Indonesia yang memiliki sejarah dukungan yang kuat terhadap perjuangan Palestina. Masyarakat di Indonesia sering kali mengekspresikan rasa solidaritas, keprihatinan, dan pendapat mereka tentang konflik tersebut melalui berbagai media, termasuk platform online dan media sosial. Salah satu media sosial, X (Twitter), berfungsi sebagai wadah publik untuk menyampaikan pendapat mereka tentang masalah tersebut. Oleh karena itu, berbagai data tweet pengguna mengenai konflik Palestina-Israel telah dikumpulkan. Data ini dapat diolah dan dianalisis untuk memahami perasaan pengguna aplikasi X (Twitter) terhadap konflik tersebut [1].

Berdasarkan penelitian terdahulu yang dilakukan oleh Upa dengan judul "Analisis Sentimen Pengguna Twitter Menggunakan Metode SVM-KNN" mengklasifikasikan sentimen pengguna X (Twitter) menggunakan gabungan

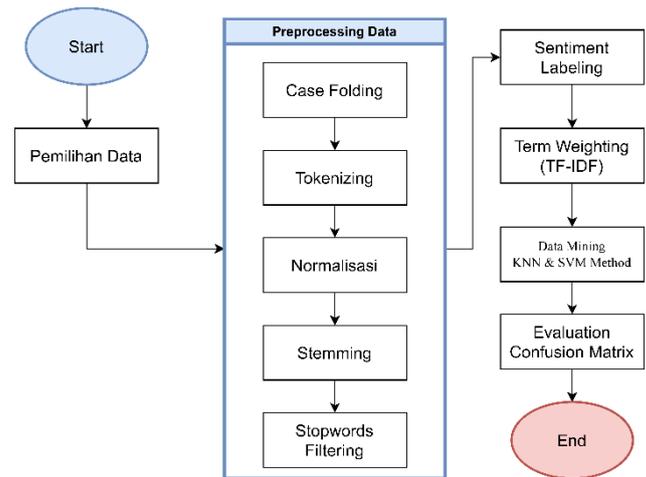
algoritma SVM dan KNN. Hasil penelitian menunjukkan bahwa pendekatan SVM-KNN memberikan kinerja yang lebih baik dibandingkan penggunaan SVM saja dalam hal presisi, recall, dan akurasi. Data diambil dari X (Twitter), menggunakan pemodelan N-Gram untuk ekstraksi fitur. Analisis menunjukkan bahwa metode gabungan SVM-KNN mencapai presisi sebesar 82%, recall 81,5%, dan akurasi 86%, yang lebih tinggi dibandingkan SVM dengan presisi 79%, recall 78%, dan akurasi 81%. Temuan ini mengindikasikan bahwa kombinasi SVM-KNN dapat meningkatkan performa klasifikasi sentimen pada data X (Twitter) [2]. Penelitian lain yang dilakukan oleh Susilawati, Lestari dan Nina mengenai sentimen masyarakat Indonesia terkait konflik Israel-Palestina, khususnya dalam konteks boikot produk Israel melalui platform media sosial X (Twitter). Lebih dari 300 dataset tweet berhasil dikumpulkan melalui metode scraping. Penelitian ini mengindikasikan bahwa sebagian besar masyarakat mendukung boikot terhadap produk Israel, dengan akurasi klasifikasi Naive Bayes mencapai 95%. Selain itu, nilai Precision, Recall, dan F1 Score masing-masing

tercatat sebesar 96%, 95%, dan 95%. Proses preprocessing data berhasil mengurangi gangguan dan mempersiapkan data untuk analisis sentimen yang lebih mendalam, menunjukkan bahwa sebagian besar pengguna yang terlibat dalam percakapan di X (Twitter) cenderung mendukung boikot produk Israel, serta menunjukkan tingkat keberhasilan yang tinggi dalam mendeteksi sentimen dari tweet yang dianalisis. [3].

Pentingnya identifikasi sentimen masyarakat Indonesia terhadap konflik di Palestina melalui platform media sosial seperti X (Twitter) tidak dapat diabaikan. Dalam konteks dinamika sosial dan politik global, pemahaman mengenai respons dan pandangan publik terhadap isu sensitif ini memiliki nilai yang sangat besar. Analisis sentimen terhadap tweet dan komentar di X (Twitter) memberikan pemahaman yang lebih dalam mengenai perasaan, pendapat, dan solidaritas masyarakat Indonesia terkait konflik tersebut, dengan membagi pandangan ke dalam beberapa kategori sentimen seperti: positif, negatif, atau netral. Metode klasifikasi yang digunakan adalah *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM). Pemilihan kedua metode ini didasarkan pada kemampuan KNN untuk memberikan hasil yang cepat dan sederhana dalam klasifikasi, serta efektivitas SVM dalam menangani data berdimensi tinggi dan menghasilkan akurasi yang tinggi dalam klasifikasi. Dengan demikian, diharapkan kedua metode ini dapat memberikan pemahaman yang lebih mendalam tentang pola sentimen serta memungkinkan perbandingan efektivitas dan akurasi masing-masing metode. Klasifikasi KNN memberikan gambaran awal mengenai sentimen, sementara SVM diharapkan meningkatkan akurasi klasifikasi. Penelitian ini berkontribusi pada pemahaman pola sentimen dan juga penting untuk menjelaskan bagaimana hasil penelitian ini dapat dimanfaatkan dalam konteks nyata; bagi pemerintah atau organisasi sosial, pemahaman mengenai opini publik dapat membantu merumuskan kebijakan yang lebih responsif terhadap kebutuhan masyarakat dan mengembangkan strategi komunikasi yang lebih efektif dalam menyikapi isu-isu sensitif seperti konflik di Palestina.

## II. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini terdiri atas enam tahapan utama yang disusun secara sistematis guna memastikan kelancaran penelitian sesuai dengan tujuan yang telah ditetapkan. Tahapan-tahapan tersebut meliputi pemilihan data, preprocessing data, pelabelan sentimen, transformasi, data mining, serta evaluasi. Setiap tahapan akan dijelaskan lebih lanjut pada bagian-bagian berikutnya.



Gambar 1. Tahapan Penelitian

Gambar 1 menunjukkan tahapan penelitian ini, terdapat 6 proses utama yang dilakukan. Langkah pertama adalah Pemilihan Data. Tahap berikutnya adalah Proses Preprocessing, Pelabelan Sentimen, Term Weighting dengan TF-IDF, Data Mining dan Evaluation dengan *Confusion Matrix*.

### A. Pemilihan Data

Tahap pengumpulan data X (Twitter) dilakukan dengan menggunakan alat bernama tweet-harvest, yang merupakan aplikasi yang dapat dijalankan menggunakan Node.js. Tweet Harvest adalah library yang digunakan untuk mengumpulkan data dari X (Twitter) dengan memanfaatkan `auth_token` X (Twitter). Metode ini memungkinkan pengambilan data berupa komentar pengguna sebagai sumber data [4]. Dalam prosesnya, beberapa kueri seperti 'Palestina', 'Israel', 'Hamas', dan 'Gaza' dimasukkan ke dalam program, dan filter bahasa serta tanggal diterapkan untuk mendapatkan data yang relevan.

### B. Preprocessing Data

Proses preprocessing data merupakan tahapan yang sangat penting yang harus dilakukan sebelum data digunakan dalam pembuatan model analisis sentimen [5]. Pada tahap ini terdapat beberapa proses seperti case folding, tokenizing, normalisasi, stemming, dan filter stopwords.

#### 1. Case Folding

Tahap Case Folding merupakan proses merubah data opini menjadi lowercase [6]. Proses ini penting untuk memastikan bahwa kata-kata yang memiliki huruf kapital dan huruf kecil diperlakukan sama dalam analisis, sehingga meningkatkan akurasi pengolahan data.

#### 2. Tokenizing

Proses tokenisasi adalah pemisahan string input menjadi bagian-bagian kecil berupa kata-kata yang menyusunnya [7]. Dengan melakukan tokenisasi, kita dapat lebih mudah mengidentifikasi makna dan konteks dari setiap kata, yang

sangat penting dalam analisis sentimen dan pemrosesan bahasa alami.

### 3. Normalisasi

Normalisasi adalah langkah yang mengubah kata-kata tidak baku menjadi bentuk baku sesuai dengan kosakata yang terdapat dalam KBBI [8]. Proses ini sangat penting untuk menjamin konsistensi data yang digunakan dalam analisis serta memastikan bahwa data tersebut memenuhi standar bahasa, sehingga meningkatkan akurasi hasil analisis yang dilakukan. Contoh kata baku dapat dilihat pada tabel 1.

TABEL I  
CONTOH KATA BAKU

Kata Tidak Baku	Kata Baku
skrng	sekarang
tdk	tidak
klo	kalau

### 4. Stemming

Tahap stemming bertujuan untuk mengembalikan kata-kata ke bentuk dasarnya sesuai dengan yang tertera di kamus. Dalam penelitian ini, proses stemming dilakukan dengan menggunakan referensi dari library Sastrawi. Melalui tahap ini, data tweet akan diproses untuk membatasi jumlah kata, sehingga setiap kata yang tersisa menjadi poin-poin penting [9].

### 5. Filter Stopwords

Filter stopwords adalah proses yang berfungsi untuk menghilangkan kata-kata yang dianggap tidak sesuai atau relevan dalam data tweet. Contoh kata stopwords yang umum meliputi “dari”, “ke”, “yang”, “di”, “dan”, serta kata-kata sejenis lainnya [10]. Dengan menghapus stopwords, analisis dapat lebih fokus pada kata-kata yang memberikan informasi penting dan makna, sehingga meningkatkan kualitas pemrosesan data.

### C. Pelabelan Sentimen

Dalam penelitian ini, pelabelan sentimen dilakukan dengan metode Lexion-based. Metode ini bergantung pada kamus opini, atau kamus, untuk menentukan klasifikasi. Dalam kamus ini, kata-kata yang digunakan untuk menggambarkan berbagai jenis pendapat, seperti positif dan negatif [11], sedangkan kata-kata yang tidak terdapat dalam kamus diberi label netral. Untuk menyeimbangkan pelabelan data, metode *Synthetic Minority Oversampling Technique* (SMOTE) diterapkan. Metode ini berfungsi dengan menghasilkan replika dari data minoritas, yang disebut data sintetis. Proses SMOTE dilakukan dengan mengidentifikasi k tetangga terdekat untuk setiap data dalam kelas minoritas. Kemudian, data sintetis dibuat sesuai dengan persentase duplikasi yang diinginkan antara data minor dan k tetangga terdekat yang dipilih secara acak [12].

### D. Term Weighting

Data teks diubah menjadi vektor numerik melalui metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Pembobotan adalah proses untuk mengubah kata masukan menjadi data numerik [13]. Untuk menghitung TF dengan menghitung setiap frekuensi kata yang muncul dalam dokumen. Sedangkan IDF menunjukkan bobot kelangkaan. Adapun Persamaan untuk menghitung IDF dapat dilihat pada persamaan 1.

$$W_{ij} = tf_{ij} \times \log\left(\frac{N}{df_j}\right) \quad (1)$$

dimana:

$W_{ij}$  = Nilai bobot TF-IDF untuk kata ke-i di dokumen j.

$tf_{ij}$  = Frekuensi kemunculan kata ke-i dalam dokumen j.

$N$  = Jumlah total dokumen dalam kumpulan data.

$df_j$  = Jumlah dokumen yang mengandung kata ke-j.

Setelah itu, dilakukan seleksi fitur menggunakan metode Chi-Square. Chi-square adalah metode yang dipilih dalam penelitian ini untuk mengurangi jumlah fitur. Fitur yang dipilih adalah fitur yang memiliki korelasi yang kuat dalam proses klasifikasi [14]. Persamaan chi-square dapat dilihat pada persamaan 2.

$$\chi^2 = \sum \frac{(Ox - Ex)^2}{2} \quad (2)$$

dimana:

$Ox$  = Frekuensi yang diamati dari fitur x.

$Ex$  = Frekuensi yang diharapkan dari fitur x.

### E. Data Mining

Penelitian ini menggunakan dua metode klasifikasi yaitu *K-Nearest Neighbors* (KNN) dan *Support Vector Machine* (SVM). Metode *K-Nearest Neighbors* (KNN) digunakan untuk mengklasifikasi sentimen berdasarkan kemiripan datanya. KNN menggunakan metrik jarak untuk menentukan kedekatan data. Pada tugas klasifikasi, KNN beroperasi dengan menemukan K titik data terdekat dalam himpunan pelatihan yang paling dekat dengan sampel yang akan diklasifikasikan, berdasarkan ukuran jarak seperti jarak Euclidean. [15]. Berikut ini rumus persamaanya :

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3)$$

dimana :

$x_i, y_i$  = nilai fitur dari dua titik data yang dibandingkan

$n$  = jumlah fitur.

Setelah menghitung jarak untuk semua titik dalam data train, pilih k tetangga terdekat (biasanya 3 atau lebih). Misalnya, jika  $k = 3$ , maka algoritma akan memilih tiga titik terdekat dengan jarak terkecil.

Sementara metode *Support Vector Machine* (SVM) adalah salah satu metode dalam kategori supervised learning. Metode ini menggunakan hipotesis yang berbentuk fungsi linear dari fitur yang memiliki dimensi tinggi. Proses pelatihan fitur dilakukan dengan algoritma pembelajaran yang berlandaskan teori optimasi. SVM dapat diterapkan untuk mengklasifikasikan dua kelas dan efektif dalam menangani baik data linear maupun non-linear [1].

F. Evaluation

Hasil klasifikasi kedua model dinilai menggunakan *Confusion Matrix*. Tahap ini juga mencakup proses evaluasi prosedur, dengan tujuan menilai seberapa efisien model melalui pengukuran Akurasi, Precision, Recall, dan F1-Score [16]. Berikut ini tabel *Confusion Matrix* dan rumus setiap hasil evaluasinya:

TABEL II  
CONFUSION MATRIX SENTIMEN

Aktual / Prediksi	Negatif	Netral	Positif
Negatif	True Negatif (TN)	False Netral (FN)	False Positif (FP)
Netral	False Negatif (FN)	True Netral (TN)	False Positif (FP)
Positif	False Negatif (FN)	False Netral (FN)	True Positif (TP)

$$Akurasi = \frac{Jumlah\ Prediksi\ Benar}{Total\ Jumlah\ Data\ Uji} \tag{4}$$

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Positif\ (FP)} \tag{5}$$

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)+False\ Negative\ (FN)} \tag{6}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

Hasil dari evaluasi ini akan memberikan pemahaman yang lebih baik tentang kemampuan setiap model dalam menangkap sentimen yang berbeda dari data yang dianalisis.

III. HASIL DAN PEMBAHASAN

A. Pemilihan Data

Opini masyarakat Indonesia terkait konflik di Palestina di media sosial X (Twitter) sangat beragam dalam menanggapi konflik tersebut. Oleh karena itu, opini-opini yang disuarakan di media sosial X (Twitter) ini akan menjadi sumber data pada penelitian. Dari tanggal 16 Mei 2024 hingga 1 Oktober 2024, sebanyak 2960 data Twitter berhasil dikumpulkan.

TABEL III  
CONTOH DATA AWAL HASIL CRAWLING

No	created_at	full_text	username
1	Mon Mar 11 23:59:09 +0000 2024	Ya Allah Selamatkan Saudara Kami #FreePalestine #CeaseFireInGaza #IsraelWarCrimes #IsraelNewNazism #Resistance #BoycottIsraeliProducts	asimazehra
2	Mon Mar 11 15:18:54 +0000 2024	Salam ya Palestine 🇵🇸🇵🇸🇵🇸 #FreePalestine https://t.co/idIg0eLg6u	fajaronline
3	Mon Mar 11 15:23:23 +0000 2024	Serangan israel di kota #gaza bagian barat. #FreePalestine	AfiyahKamil a7

B. Preprocessing Data

Tahap berikutnya adalah preprocessing data, yang merupakan langkah penting dalam proses analisis sentimen. Contoh hasil dari tahapan preprocessing dapat dilihat pada tabel 4.

TABEL IV  
CONTOH HASIL PERPROCESSING DATA

Tahap Preprocessing	Hasil
Data Asli	Serangan Israel di kota bagian barat
Case Folding	serangan israel di kota bagian barat
Tokenizing	['serangan', 'israel', 'di', 'kota', 'bagian', 'barat']
Normalisasi	['serangan', 'israel', 'di', 'kota', 'bagian', 'barat']
Stemming	['serang', 'israel', 'di', 'kota', 'bagi', 'barat']
Filter Stopwords	['serang', 'israel', 'kota', 'barat']

Hasil preprocessing yang ditampilkan dalam Tabel 4 menunjukkan transformasi dari data asli hingga menjadi data yang lebih bersih dan terstruktur. Setiap tahap, mulai dari case folding hingga filter stopwords, berkontribusi untuk meningkatkan kualitas data yang akan digunakan dalam analisis sentimen.

C. Pelabelan Sentimen

Dalam Proses pelabelan sentimen dilakukan dengan menggunakan metode Lexion-based. Setiap data dikategorikan sebagai positif, negatif, atau netral. Untuk melakukan hal ini, pendekatan Lexion-based memanfaatkan kamus sentimen yang berisi kata-kata positif dan negatif. Jika sebuah kata tidak ditemukan dalam kamus tersebut, maka label netral diberikan. Selain itu, metode SMOTE juga

diterapkan untuk menyeimbangkan data. Contoh dataset hasil pelabelan dapat dilihat pada tabel 5.

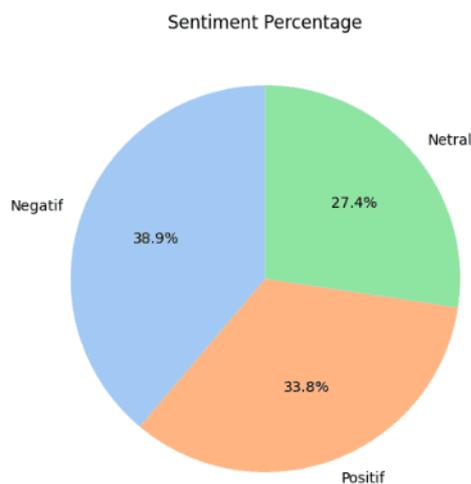
TABEL V  
CONTOH HASIL PELABELAN SENTIMEN

No	Text Clean	Label
1	selamat saudara	Positif
2	salam palestine	Netral
3	doa saudara saudara palestina moga senantiasa kuat tabah sabar berkah palestine tomorrow will be free	Positif

TABEL VI  
JUMLAH DATA

No	Label	Jumlah
1	Negatif	1150
2	Positif	1000
3	Netral	810
	<b>Jumlah</b>	2960

Tabel 6 menunjukkan distribusi data sentimen, yaitu 1150 untuk sentimen negatif, 1000 untuk sentimen positif, dan 810 untuk sentimen netral, dengan total 2960 data.



Gambar 2. Presentase Sentimen

Gambar 2 menunjukkan persentase distribusi data sentimen. Sentimen negatif memiliki porsi terbesar dengan 38.9%, diikuti oleh sentimen positif sebesar 33.8%, dan sentimen netral sebanyak 27.4%. Dominasi sentimen negatif ini dapat dipahami dalam konteks konflik di Palestina, di mana banyak pengguna media sosial menyampaikan opini yang menjelek-jelekan atau menghina pihak Israel, mengaitkan tindakan mereka dengan penderitaan yang dialami oleh warga Palestina. Di sisi lain, sentimen positif muncul sebagai bentuk dukungan terhadap perjuangan rakyat Palestina dan upaya mencapai keadilan. Sementara itu, sentimen netral sebagian besar berisi kalimat doa dan harapan untuk perdamaian, mencerminkan keinginan masyarakat

untuk melihat resolusi damai terhadap konflik yang berkepanjangan ini.

#### D. Term Weighting

Setelah proses pelabelan selesai, proses pembobotan kata atau term weighting menggunakan TF-IDF dilakukan. Contoh hasil term weighting dapat dilihat pada tabel 7.

TABEL VII  
CONTOH HASIL TERM WEIGHTING DENGAN TF-IDF

tweet/ komentar	Kata				
	1	2	3	4	...
1	0.00	0.00	0.00	0.00	...
2	0.85	0.00	0.00	0.00	...
3	0.00	0.00	0.26	0.00	...
4	0.37	0.00	0.40	0.00	...
...	...	...	...	...	...
2.960	0.00	0.00	0.00	0.00	...

#### E. Data Mining

Proses data mining dilakukan dengan menggunakan metode klasifikasi K-Nearest Neighbors (KNN) secara terpisah, serta kombinasi KNN dengan Support Vector Machine (SVM). Sebelum melakukan proses klasifikasi, dataset dibagi terlebih dahulu menjadi dua, yaitu data latih dan data uji dengan rasio 80:20. Dalam penelitian ini, nilai k yang dipilih berdasarkan skor Chi-Square adalah 275. Tujuan penerapan kedua metode ini adalah untuk membandingkan efektivitas masing-masing dalam mengklasifikasikan sentimen, sehingga dapat ditentukan metode mana yang memberikan hasil lebih baik dalam konteks analisis data yang dilakukan.

TABEL VIII  
AKURASI MODEL KNN

Nilai k	Akurasi	
	KNN	KNN + SVM
3	75%	81%
5	74%	80%
7	73%	82%
9	74%	80%

Hasil eksperimen pada tabel 8 menunjukkan perbandingan antara metode *K-Nearest Neighbors* (KNN) dan gabungan *K-Nearest Neighbors* (KNN) dengan *Support Vector Machine* (SVM). Pada nilai  $k = 7$ , akurasi KNN tercatat sebesar 73%, sedangkan KNN + SVM mencapai akurasi 82%. Perbedaan ini mencerminkan efektivitas penggabungan kedua metode dalam klasifikasi sentimen. SVM memiliki keunggulan dalam menangani data yang lebih kompleks, yang memungkinkan model untuk menemukan pola yang lebih baik di antara respons sentimen. Dengan demikian, meskipun KNN memberikan hasil yang wajar, kombinasi KNN dengan SVM menunjukkan kemampuan yang lebih unggul dalam menangkap nuansa sentimen masyarakat terkait isu-isu sensitif, seperti konflik di Palestina.

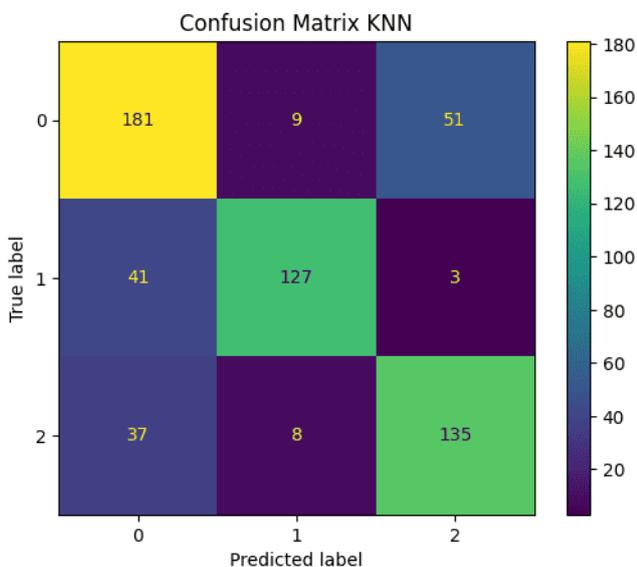
F. Evaluation

Berikut ini adalah tabel hasil evaluasi analisis sentimen yang telah dilakukan, yang menunjukkan nilai dari berbagai evaluasi seperti Akurasi, Precision, Recall, dan F1-Score untuk setiap nilai metode yang digunakan.

TABEL IX  
EVALUASI HASIL ANALISIS

KNN				
k	Akurasi	Precision	Recall	F1-Score
3	75%	76%	75%	75%
5	74%	75%	74%	74%
7	73%	74%	73%	73%
9	74%	76%	74%	75%
KNN + SVM				
3	81%	81%	81%	81%
5	80%	81%	80%	80%
7	82%	83%	82%	82%
9	80%	80%	80%	80%

Hasil eksperimen menunjukkan bahwa model KNN + SVM dengan nilai k = 7 memberikan performa tertinggi, mencapai akurasi 82%. Pada konfigurasi ini, model juga menunjukkan tingkat Precision sebesar 83%, Recall sebesar 82%, dan F1-score sebesar 82%, yang mencerminkan keseimbangan tinggi dalam akurasi serta kemampuan klasifikasi untuk berbagai sentimen secara konsisten.

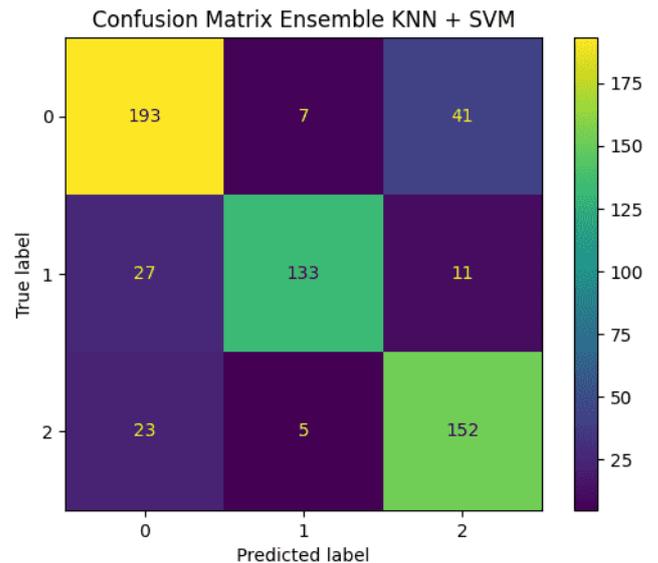


Gambar 3. Confusion Matrix KNN

Sebagai perbandingan, performa terbaik dari model KNN sendiri tercapai pada nilai k = 3, dengan akurasi 75%, Precision 76%, Recall 75%, dan F1-score 75%. Meskipun hasil ini menunjukkan bahwa KNN dapat mengklasifikasikan data dengan cukup baik, peningkatan yang signifikan terlihat ketika KNN digabungkan dengan SVM, memberikan akurasi dan efisiensi klasifikasi yang lebih baik dalam menangkap pola-pola sentimen secara lebih akurat. Berikut adalah

Confusion Matrix yang memberikan gambaran lebih lanjut tentang performa model.

Berdasarkan Confusion Matrix, model KNN mengklasifikasikan 181 sampel dari kelas 0 (positif) dengan benar, namun ada kesalahan ke kelas negatif (51). Untuk kelas 1 (netral), 127 sampel diklasifikasikan benar, dengan kesalahan terutama ke kelas positif (41). Pada kelas 2 (negatif), ada 135 prediksi benar, namun 37 sampel salah sebagai positif. Meskipun KNN cukup akurat, kesalahan lebih sering terjadi antara kelas positif dan negatif.



Gambar 4. Confusion Matrix KNN + SVM

Sementara berdasarkan Confusion Matrix KNN + SVM pada Gambar 4, gabungan kedua metode berhasil mengklasifikasikan sebagian besar sampel dengan benar di masing-masing kelas. Untuk kelas 0 (positif), terdapat 193 prediksi benar, dengan beberapa kesalahan ke kelas netral (7) dan negatif (41). Pada kelas 1 (netral), 133 sampel diklasifikasikan dengan benar, meskipun ada sedikit kesalahan ke kelas positif (27) dan negatif (11). Pada kelas 2 (negatif), 152 sampel benar diklasifikasikan, namun masih ada kesalahan ke kelas positif (23) dan netral (5). Secara keseluruhan, performa model cukup baik, meski kesalahan masih terjadi antara kelas positif dan negatif.



## DAFTAR PUSTAKA

- [1] H. Ramanizar, A. Fajri, R. Binsar Sinaga, H. Mubarak, A. D. Pangestu, and D. S. Prasvita, *Analisis Sentimen Pengguna Twitter terhadap Konflik antara Palestina dan Israel Menggunakan Metode Naïve Bayesian Classification dan Support Vector Machine*. 2021.
- [2] F. Upa and Nurhalifah, "Analisis Sentimen Pengguna Twiter Menggunakan Metode SVM-KNN Analysis of Twiter User Sentiment Using SVM-KNN Method," *Nusantara Hasana Journal*, vol. 3, no. 12, p. Page, 2024.
- [3] A. Tiara Susilawati, N. Anjeni Lestari, and P. Alpria Nina, "Analisis Sentimen Publik Pada Twitter Terhadap Boikot Produk Israel Menggunakan Metode Naïve Bayes," *Jurnal Ilmiah Mahasiswa*, vol. 2, no. 1, pp. 26–35, 2024, doi: 10.59603/niantanasikka.v2i1.240.
- [4] B. A. Yuniarossy *et al.*, "Analisis Sentimen Terhadap Isu Feminisme Di Twitter Menggunakan Model Convolutional Neural Network (CNN)," vol. 5, no. 1, 2024, doi: 10.46306/lb.v5i1.
- [5] A. Hermawan, I. Jowensen, J. Junaedi, and Edy, "Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine," *JST (Jurnal Sains dan Teknologi)*, vol. 12, no. 1, pp. 129–137, Apr. 2023, doi: 10.23887/jstundiksha.v12i1.52358.
- [6] P. Pasek, O. Mahawardana, G. A. Sasmita, P. Agus, and E. Pratama, "Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap 'Figure Pemimpin' Menggunakan Python," 2022.
- [7] T. Krisdiyanto, E. Maricha, and O. Nurharyanto, "Analisis Sentimen Opini Masyarakat Indonesia Terhadap Kebijakan PPKM pada Media Sosial Twitter Menggunakan Naïve Bayes Clasifiers," *Jurnal CoreIT*, vol. 7, no. 1, 2021.
- [8] J. W. Iskandar and Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, pp. 1120–1126, Dec. 2021, doi: 10.29207/resti.v5i6.3588.
- [9] M. G. Andriawan and T. Ernawati, "Penggunaan Algoritma Naïve Bayes Dan Support Vector Machine Untuk Analisis Sentimen Konflik Palestina Dan Israel Pada Platform X," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4943.
- [10] T. Safitri, Y. Umidah, and I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap BTS Menggunakan Algoritma Support Vector Machine," 2023. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [11] D. Musfiroh *et al.*, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, pp. 24–33, 2021.
- [12] R. Siringoringo, "Klasifikasi Data Tidak Seimbang Menggunakan Algoritma Smote dan K-Nearest Neighbor," 2018.
- [13] I. Prayoga, M. D. Purbolaksone, and A. Adiwijaya, "Sentiment Analysis on Indonesian Movie Review Using KNN Method With the Implementation of Chi-Square Feature Selection," *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, p. 369, Jan. 2023, doi: 10.30865/mib.v7i1.5522.
- [14] M. B. Hamzah, "Classification of Movie Review Sentiment Analysis Using Chi-Square and Multinomial Naïve Bayes with Adaptive Boosting," *Journal of Advances in Information Systems and Technology*, vol. 3, no. 1, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/jaist>
- [15] S. D. Prasetyo, S. S. Hilabi, and F. Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *Jurnal KomtekInfo*, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [16] R. I. Putra Selian, A. V. Vitianingsih, S. Kacung, A. Lidya Maukar, and J. Febrian Rusdi, "Sentiment Analysis of Public Responses on Social Media to Satire Joke Using Naive Bayes and KNN," *sinkron*, vol. 8, no. 3, pp. 1443–1451, Jul. 2024, doi: 10.33395/sinkron.v8i3.13721.
- [17] F. M. Carina, Admi Salma, Dony Permana, and Zamahsary Martha, "Sentiment Analysis of X Application Users on the Conflict Between Israel and Palestine Using Support Vector Machine Algorithm," *UNP Journal of Statistics and Data Science*, vol. 2, no. 2, pp. 204–212, May 2024, doi: 10.24036/ujsds/vol2-iss2/170.
- [18] D. Vonega, A. Fadila, and D. Kurniawan, "Analisis Sentimen Twitter Terhadap Opini Publik Atas Isu Pencalonan Puan Maharani dalam PILPRES 2024," *JAIC*, vol. 6, no. 2, pp. 129–135, Nov. 2022.
- [19] J. Friadi, and D. E. Kurniawan, "Analisis Sentimen Ulasan Wisatawan Terhadap Alun-Alun Kota Batam: Perbandingan Kinerja Metode Naive Bayes dan Support Vector Machine," *Jurnal Sistem Informasi Bisnis*, vol. 14, no. 4, pp. 403–407, Oct. 2024. <https://doi.org/10.21456/vol14iss4pp403-407>
- [20] S. D. Widiyaningsih and A. Pertiwi, "Analysis Of Ovo Application Sentiment Using Lexicon Based Method And K-Nearest Neighbor," *Jurnal Ilmiah Ekonomi Bisnis*, vol. 25, no. 1, pp. 14–28, 2020, doi: 10.35760/eb.2020.v25i1.2416.