

Comparison of KNN and SVM Performance in 2024 Election Results Sentiment Analysis

M Iqbal Fahilla Bukit^{1*}, Andre Hasudungan Lubis²

¹ Teknik Informatika, Teknik, Universitas Medan Area, Medan, Indonesia

Informasi Artikel

Diterima : 27 Agustus 2025
Revisi : 6 September 2025
Publikasi : 30 September 2025

Kata Kunci:

Analisis Sentimen
Pemilu 2024
KNN
SVM
TF-IDF

ABSTRAK

Penelitian ini melakukan perbandingan kinerja algoritma K-Nearest Neighbor (KNN) dan Support Vector Machine (SVM) dalam analisis sentimen terkait hasil Pemilu 2024 dengan memanfaatkan data dari media sosial. Dataset yang digunakan berjumlah 506 entri opini publik yang dikategorikan ke dalam tiga label sentimen, yaitu positif, negatif, dan netral. Proses pengolahan data mencakup tahapan preprocessing berupa case folding, tokenisasi, penghapusan stopword, serta stemming, kemudian direpresentasikan dengan metode Term Frequency–Inverse Document Frequency (TF-IDF). Hasil pengujian memperlihatkan bahwa kedua algoritma mampu melakukan klasifikasi dengan akurasi lebih dari 70%. Algoritma KNN menghasilkan akurasi 75,49%, presisi 71,36%, recall 75,49%, dan F1-score 72,88%, sedangkan algoritma SVM menunjukkan performa yang sedikit lebih baik dengan akurasi 77,45%, presisi 70,59%, recall 77,45%, dan F1-score 72,15%. Berdasarkan analisis confusion matrix, kedua model memiliki kemampuan tinggi dalam mengklasifikasikan sentimen positif, namun masih menghadapi kendala dalam mengenali sentimen negatif dan netral akibat ketidakseimbangan distribusi data. Secara keseluruhan, penelitian ini mengindikasikan bahwa SVM lebih sesuai digunakan untuk analisis sentimen pemilu pada data teks berdimensi tinggi.

ABSTRACT

This study compares the performance of the K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms in sentiment analysis related to the 2024 election results using data from social media. The dataset used consists of 506 public opinion entries categorized into three sentiment labels: positive, negative, and neutral. The data processing involved preprocessing steps such as case folding, tokenization, stopword removal, and stemming, then represented using the Term Frequency–Inverse Document Frequency (TF-IDF) method. The test results showed that both algorithms were able to classify with an accuracy of over 70%. The KNN algorithm produced an accuracy of 75.49%, precision of 71.36%, recall of 75.49%, and an F1-score of 72.88%, while the SVM algorithm showed slightly better performance with an accuracy of 77.45%, precision of 70.59%, recall of 77.45%, and F1-score of 72.15%. Based on the confusion matrix analysis, both models have a high ability to classify positive sentiments, but still face obstacles in recognizing negative and neutral sentiments due to the imbalance in data distribution. Overall, this study indicates that SVM is more suitable for election sentiment analysis on high-dimensional text data.

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



*Penulis Koresponden

Email: ifahilla960@gmail.com

Cara sitasi IEEE:

M. I. F. Bukit & A. H. Lubis, "Comparison of KNN and SVM Performance in 2024 Election Results Sentiment Analysis," *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 3, pp. 1166-1173, September 2025, doi: 10.30811/jaise.v5i3.7659

1. PENDAHULUAN

Pemilihan umum (pemilu) merupakan salah satu peristiwa politik pada sistem pemerintahan demokrasi paling krusial yang menentukan arah kebijakan dan masa depan suatu negara [1]. Pada era digital, opini publik semakin banyak disuarakan melalui media sosial, menjadikannya sumber data penting untuk memahami persepsi masyarakat secara real-time. Analisis sentimen berperan dalam menggambarkan dukungan, ketidakpuasan, maupun sikap netral masyarakat terhadap kandidat atau isu politik, yang berpotensi memengaruhi hasil pemilu.

Tantangan utama dalam analisis sentimen pemilu terletak pada besarnya volume data, keragaman bahasa, serta cepatnya perubahan opini publik. Pemilihan algoritma klasifikasi yang tepat menjadi kunci dalam menghasilkan prediksi yang akurat dan andal. K-Nearest Neighbor (KNN) dikenal sederhana dan efektif dalam menangani data non-linear, sedangkan Support Vector Machine (SVM) unggul dalam memisahkan kelas dengan margin optimal serta lebih tahan terhadap noise [2].

Berbagai penelitian sebelumnya menunjukkan kecenderungan SVM memiliki performa lebih unggul dibandingkan KNN. Misalnya, studi pada Pemilu Presiden Indonesia 2019 melaporkan akurasi rata-rata SVM sebesar 69,27%, lebih tinggi dibandingkan KNN dengan 61,3% [3]. Penelitian lain pada berita Myanmar juga menunjukkan dominasi SVM dalam mengklasifikasikan sentimen positif dan negatif [1]. Hasil serupa diperoleh pada analisis data Twitter dan berbagai dataset lain, di mana SVM konsisten lebih baik dalam hal akurasi, presisi, recall, dan F1-score [4][5].

Penelitian terbaru yang memperlihatkan keunggulan pada metode SVM dibandingkan KNN dalam analisis sentimen politik. Penelitian oleh [6] menunjukkan bahwa SVM memiliki akurasi lebih stabil dibandingkan KNN pada analisis sentiment terkait pasangan calon presiden 2024. Penelitian lain oleh yang meneliti sentimen publik terhadap hasil Pemilu 2024 di platform X juga menemukan bahwa KNN memiliki akurasi lebih rendah dibandingkan algoritma lain [7]. [8] Sementara itu, menemukan bahwa SVM memberikan performa akurasi terbaik (79,57%) pada analisis sentimen Pilpres Indonesia jika dibandingkan dengan KNN maupun Naïve Bayes.

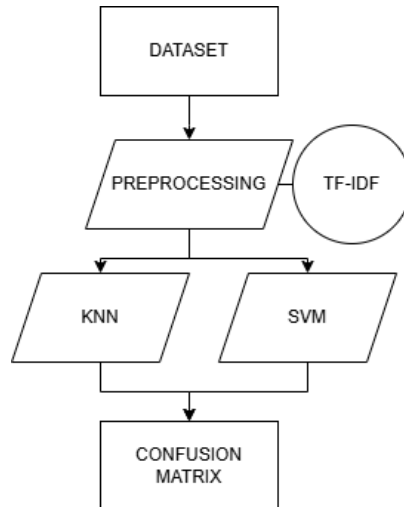
Meskipun demikian, terdapat keterbatasan pada studi-studi terdahulu, seperti penggunaan dataset yang terbatas, kurangnya pemanfaatan fitur modern (misalnya word embedding), serta minimnya penelitian yang relevan dengan Pemilu 2024 [9]. Selain itu, sebagian besar penelitian hanya fokus pada dua kelas sentimen (positif dan negatif), padahal dalam konteks pemilu, kategori netral dan ambivalen juga memegang peran penting [10].

Untuk mengisi celah tersebut, penelitian ini mengusulkan evaluasi komprehensif terhadap performa KNN dan SVM pada analisis sentimen Pemilu 2024 dengan memanfaatkan data terbaru dari media sosial. Pendekatan ini akan mengintegrasikan metode feature extraction modern seperti TF-IDF dan word embedding (Word2Vec), serta mempertimbangkan aspek akurasi, kecepatan, ketahanan terhadap noise, dan kemampuan generalisasi algoritma [11].

Kebaruan penelitian ini terletak pada pemanfaatan data real-time Pemilu 2024, penerapan teknik representasi fitur yang lebih mutakhir, serta evaluasi komparatif terhadap KNN dan SVM dalam konteks politik Indonesia yang dinamis. Diharapkan hasil penelitian ini mampu memberikan kontribusi signifikan bagi pengembangan metode analisis sentimen yang lebih akurat dan adaptif, sekaligus mendukung proses prediksi hasil pemilu dan pengambilan keputusan strategis oleh pemangku kepentingan [12].

2. METODE

Langkah-langkah penelitian divisualisasikan dengan gambar 1 alur pengambilan data



Gambar 1. Alur pengambilan data

Setiap tahapan pada alur pengambilan data adalah sebagai berikut :

2.1. Pengambilan Data

Data dikumpulkan dari Kaggle dari link ini <https://www.kaggle.com/datasets/enjelita/data-tugas-akhir/data>. Kata kunci pencarian pemilu twitter dan penulis hanya mengambil dataset dari Hasil Pemilu dari banyaknya dataset yang sediakan, data ini digunakan untuk mengumpulkan tweet yang relevan dengan subjek penelitian. Informasi yang penulis dapatkan berupa teks tweet dari hasil sentimen pemilu pada tahun 2024 dan kemudian disimpan dalam format csv agar lebih mudah diproses.

2.2. Preprocessing Data

Dataset yang digunakan melalui kaggle terdiri dari 506 data teks mengenai opini masyarakat, lalu memasuki tahapan *Preprocessing* yaitu case folding, tokenisasi, stopword, removal, dan stemming, dan menghasilkan dua kolom utama, yaitu teks bersih *clean* dan label sentimen *predicted_label* dengan tiga kelas, positif, negatif, dan netral.

Representasi teks dilakukan dengan Term Frequency-Inverse Document Frequency (TF-IDF) dengan bata maksimal 5000 fitur. Teknik ini dipilih karena mampu merepresentasikan bobot kata secara proporsional terhadap munculnya di seluruh korpus.

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

$$IDF(t) = \log \frac{N}{df_t} \quad (2)$$

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

Keterangan :

$f_{t,d}$ = frekuensi kemunculan kata t pada dokumen d

$\sum_k f_{k,d}$ = total kata dalam dokumen d

N = jumlah total dokumen

df_t = jumlah dokumen yang mengandung kata t

2.3. Pemodelan

K-Nearest Neighbor (KNN) digunakan dengan jumlah k awal yaitu $k=5$, kemudian akan dioptimasi dengan beberapa nilai k (3,5,7,9) dengan menggunakan jarak antar data dihitung dengan rumus euclidean sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

Sedangkan pada Support Vector Machine (SVM) digunakan dengan karnel yang linier, sesuai dengan data teks berdimensi tinggi yang bekerja dengan mencari hyperplane optimal yang memisahkan kelas dengan margin maksimum. Fungsi klasifikasinya dirumuskan dengan :

$$f(x) = \text{sign}(w \cdot x + b) \quad (5)$$

Dengan :

w = vektor bobot,

x = vektor input,

b = bias.

2.4. Confusion Matrix

Evaluasi model untuk membandingkan performa KNN dan SVM menggunakan confusion matrix yang memakai empat metrik utama, yaitu akurasi, presisi, recall, dan F1-score yang menggunakan rumus seperti :

Akurasi ialah mengukur sejauh mana model dapat mengklasifikasikan semua kelas dengan benar

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

Presisi yaitu presisi untuk setiap kelas yang dihitung dengan rumus berikut,

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Recall untuk setiap kelas dihitung menggunakan rumus,

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

F1-score adalah harmoni dari presisi dan recall, dan dapat dihitung dengan rumus berikut,

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

keterangan :

TP = True Positif (jumlah prediksi benar untuk kelas positif)

TN = True Negatif (jumlah prediksi benar untuk kelas negatif)

FP = False Positif (jumlah prediksi salah untuk kelas positif)

FN = False Negatif (jumlah prediksi salah untuk kelas positif)

3. HASIL DAN PEMBAHASAN

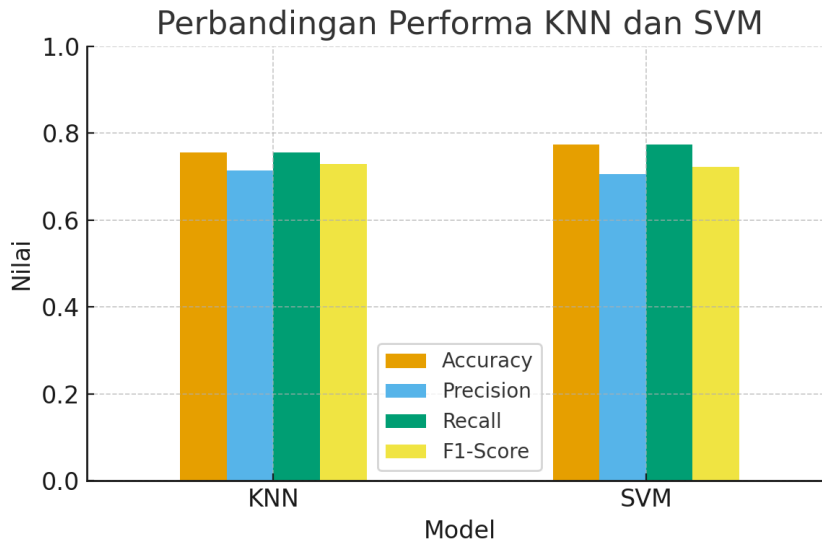
3.1. Hasil Evaluasi

Model KNN ($k=5$) dan SVM dilatih dengan data latih 80% dan diuji pada data uji 20%. Hasil evaluasi menunjukkan bahwa algoritma SVM akan lebih unggul dibandingkan KNN pada seluruh metrik dijelaskan pada *tabel 1*.

Tabel 1. Ringkasan perbandingan performa antara KNN dan SVM

Model	Accuracy	Precision	Recall	F1-Score
KNN	0.754902	0.713687	0.754902	0.728845
SVM	0.774510	0.705882	0.774510	0.721538

Hasil evaluasi menunjukkan bahwa baik algoritma KNN ataupun SVM mampu melakukan klasifikasi sentimen hasil pemilu dengan rata-rata tingkat akurasi di atas 70%. Terlihat pada bagian tabel 1 terlihat bahwa SVM memperoleh hasil akurasi yang jika dipersentasekan berjumlah sebesar 77,45%, lebih tinggi dibandingkan KNN dengan akurasi 75,49%. Selain itu, nilai untuk reccall SVM 77,45% yang juga lebih tinggi dibandingkan KNN 75,49%, hal ini menunjukkan bahwa SVM lebih baik dalam mendeteksi data positif secara konsisten dan juga sesuai dengan karakteristik SVM yang efektif dalam mengatasi data berdimensi tinggi dengan margin pemisah yang optimal.



Gambar 1. Grafik perbandingan performa

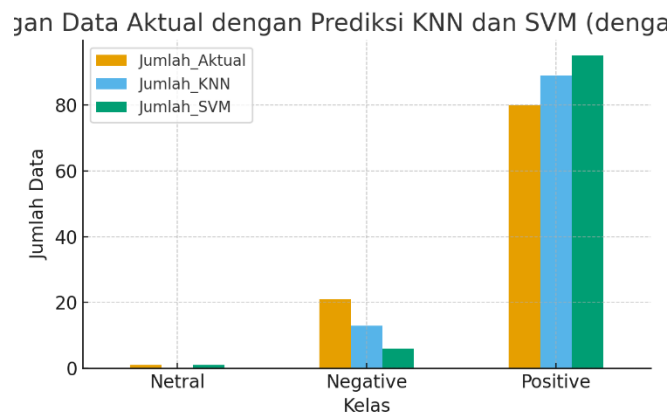
3.2. Hasil Eksperimen

Model KNN dan SVM penulis melakukan eksperimen dengan membandingkan dengan data aktual berdasarkan dataset yang ada pada kolom predicted label yaitu positif, negatif, dan netral dapat dilihat pada tabel 2.

Tabel 2. Perbandingan kualifikasi metode dengan Data aktual

Kelas	Jumlah Aktual	Prediksi KNN	Prediksi SVM
Positif	80	89	95
Negatif	21	13	6
Netral	1	0	1

Hasil klasifikasi menunjukkan bahwa kelas **positif** memiliki recall yang tinggi pada kedua model yaitu 90% untuk KNN dan 95% untuk SVM. Namun, pada kelas **negatif** masih sulit dikenali dengan baik, ditunjukkan, pada rendahnya recall yaitu 24% untuk KNN dan 10% untuk SVM.



Gambar 2. Perbandingan Data Aktual dengan prediksi KNN dan SVM

3.3. Hasil Klasifikasi

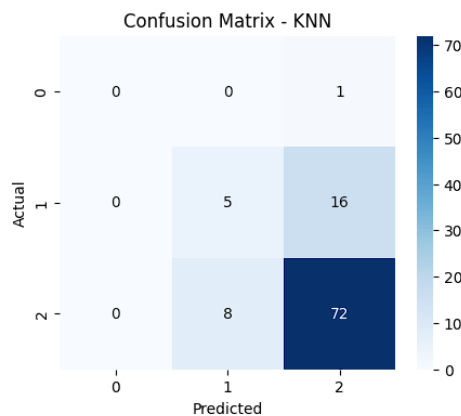
Pada tahap ini, telah dilakukan membuat model yang diuji dengan menggunakan *data training* dan *data test* yang diambil secara acak dari keseluruhan dataset.[7]

classification report				
	precision	recall	f1-score	support
nan	0.00	0.00	0.00	1
negative	0.38	0.24	0.29	21
positive	0.81	0.90	0.85	80
accuracy			0.75	102
macro avg	0.40	0.38	0.38	102
weighted avg	0.71	0.75	0.73	102

Gambar 3. Hasil Evaluasi KNN

Klasifikasi menggunakan K-Nearest Neighbor memperoleh akurasi sebesar 75%, presisi 71%, recall 75%, dan f1-score 73%. Proses ini ialah mencakup data latih dan uji, ekstraksi fitur, pelatihan model, evaluasi model, dan pembuatan *confusion matrix*. Diatas ialah hasil evaluasi dengan menggunakan model K-Nearest Neighbor Gambar 3.

Pada confusion matrix model juga memperoleh hasil akurasi sebesar 75%, presisi 71%, recall 75%, dan f1-score 73% Gambar 4.



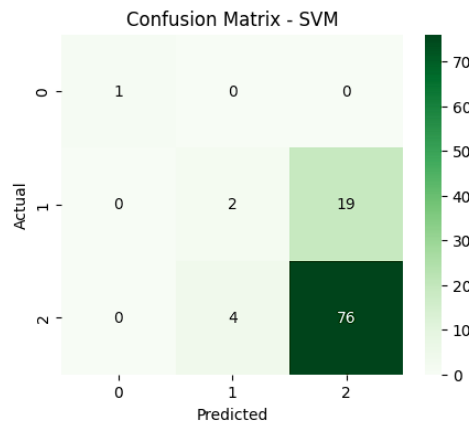
Gambar 4. Confusin matrix KNN

Sedangkan untuk klasifikasi menggunakan SVM memperoleh akurasi sebesar 77%, presisi 71%, recall 77%, dan f1-score 72%. Prosesnya sama seperti dengan evaluasi KNN yaitu mencakup data latih dan uji, ekstraksi fitur, pelatihan model, evaluasi model, dan pembuatan *confusion matrix* dapat dilihat Gambar 5.

classification report				
	precision	recall	f1-score	support
nan	1.00	1.00	1.00	1
negative	0.33	0.10	0.15	21
positive	0.80	0.95	0.87	80
accuracy			0.77	102
macro avg	0.71	0.68	0.67	102
weighted avg	0.71	0.77	0.72	102

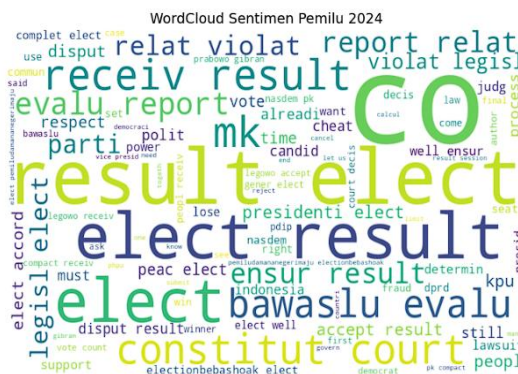
Gambar 5. Hasil Evaluasi model SVM

Pada confusion matrix model SVM, hasil klasifikasinya memperoleh akurasi sebesar 77%, presisi 71%, recall 77%, dan f1-score 72% terlihat pada Gambar 6.



Gambar 6. Confusion Matrix SVM

Hasil klasifikasi analisis sentimen dengan keyword "Hasil Pemilu 2024" juga dapat dilihat dengan visualisasi wordcloud sebagai berikut Gambar 7.



Gambar 7. Hasil Wordcloud

3.4. Pembahasan

Hasil menunjukkan bahwa SVM memberikan kinerja yang baik dibandingkan KNN, walaupun selisih diantaranya relatif kecil. Hal ini tepat dengan teori bahwa SVM lebih efektif dalam menangani data berdimensi tinggi dan non-linier.

Namun, model KNN dan SVM memiliki kelemahan dalam mendeteksi sentimen negatif. Salah satu penyebabnya ialah faktor distribusi data yang tidak seimbang, di mana kelas positif lebih dominan daripada kelas negatif. Pada kondisi ini membuat model yang cenderung bias terhadap prediksi positif, sehingga mengurangi kemampuan generalisasi pada kelas minoritas. Adanya kelas netral juga menunjukkan bahwa model masih terbatas dalam mengenali kategori secara konsisten

Perspektif analisis sentimen politik, hal ini dapat memiliki implikasi penting. Kesalahan dalam mengenali opini negatif ataupun netral dapat mengurangi validitas analisis terhadap dinamika persepsi publik. Oleh karena itu pendekatan misalnya seperti *oversampling*(SMOTE) atau *cost-sensitive learning* yang dapat digunakan untuk penelitian selanjutnya agar adanya keseimbangan data.

Hasil penelitian ini menunjukkan bahwa SVM lebih direkomendasikan untuk analisis sentimen hasil pemilu 2024 dibandingkan KNN, utamanya pada konteks data teks kompleks dan berdimensi tinggi.

4. KESIMPULAN

Penelitian ini membandingkan kinerja algoritma KNN dan SVM dalam menganalisis sentimen hasil Pemilu 2024. Hasil menunjukkan bahwa kedua algoritma dapat mengklasifikasikan sentimen dengan akurasi di atas 70%. SVM sedikit lebih baik daripada KNN, mencapai akurasi 77,45% dibandingkan dengan 75,49%. Kedua metode ini cukup baik dalam mengidentifikasi sentimen positif, tetapi mengalami kesulitan dalam mengidentifikasi sentimen negatif dan netral karena datanya tidak seimbang.

Hasil ini menunjukkan bahwa SVM merupakan pilihan yang baik untuk menganalisis sentimen berbasis teks, terutama dalam situasi politik yang berubah dengan cepat seperti Pemilu 2024.

Untuk meningkatkan kualitas, penelitian di masa mendatang sebaiknya mencakup teknik seperti SMOTE atau pembelajaran yang sensitif terhadap biaya untuk menyeimbangkan data, dan menggunakan cara yang lebih baru untuk merepresentasikan kata, seperti penyisipan kata, guna meningkatkan akurasi, terutama untuk jenis sentimen yang kurang umum.

REFERENSI

- [1] T. Yu and K. Nwet, "Comparing SVM and KNN Algorithms for Myanmar News Sentiment Analysis System," *Proc. 2020 6th Int. Conf. Comput. Data Eng.*, 2020, doi: 10.1145/3379247.3379293.
- [2] G. Dlamini, Z. Kholmatova, A. Kruglov, G. Succi, H. Tarasau, and A. Valeev, "Meta-analytical Comparison Of SVM and KNN for Text Classification," *2021 Int. Conf. "Nonlinearity, Inf. Robot.*, pp. 1–6, 2021, doi: 10.1109/NIR52917.2021.9666133.
- [3] F. Firmansyah *et al.*, "Comparing Sentiment Analysis of Indonesian Presidential Election 2019 with Support Vector Machine and K-Nearest Neighbor Algorithm," *2020 6th Int. Conf. Comput. Eng. Des.*, pp. 1–6, 2020, doi: 10.1109/ICCED51276.2020.9415767.
- [4] S. Al Hasan, A. Noman, and J. Ji, "A Performance Evaluation of Sentiment Classification Applying SVM, KNN, and Naive Bayes," *2021 Int. Conf. Comput. Networking, Telecommun. Eng. Sci. Appl.*, pp. 56–60, 2021, doi: 10.1109/contesa52813.2021.9657115.
- [5] A. B. Altinel, "Türkçe Metinlerde Makine Öğrenmesi Algoritmalarının Duygu Analizi Problemi Üzerindeki Performansının Kıyaslanması," *Eur. J. Sci. Technol.*, 2021, doi: 10.31590/ejosat.1011864.
- [6] Muhammad Hanafi and Mhd.Furqan, "Perbandingan Analisis Sentimen Presiden 2024 Menggunakan Algoritma Support Vector Machine dan K-Nearest Neighbor," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 10, no. 1 SE-Articles, pp. 275–285, Jan. 2025, doi: 10.24114/cess.v10i1.65928.
- [7] T. Aurelly Claudia Budianto, H. Fatoni, M. Ayu Syaharani, and C. Rozikin, "Analisis Sentimen Pengumuman Hasil Pemilu 2024 Di Sosial Media X Menggunakan Knn Dan Naive Bayes Classifier," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10816–10822, 2024, doi: 10.36040/jati.v8i5.11077.
- [8] A. L. Hananto, A. P. Nardilasari, A. Fauzi, A. Hananto, B. Priyatna, and A. Y. Rahman, "Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on Twitter," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 6s, pp. 473–481, 2023.
- [9] S. H. Hasanah, M. R. Maulana, and D. Nurdiana, "GOJEK DATA ANALYSIS THROUGH TEXT MINING USING SUPPORT VECTOR MACHINE (SVM) AND K-NEAREST NEIGHBOR (KNN)," *BAREKENG J. Ilmu Mat. dan Terap.*, 2025, doi: 10.30598/barekengvol19iss2pp889-902.
- [10] Q. Zhao, "Social emotion classification of Japanese text information based on SVM and KNN," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–12, 2021, doi: 10.1007/S12652-021-03034-X.
- [11] B. Paul, S. Guchhait, T. Dey, D. Das Adhikary, and S. Bera, "A Comparative Study on Sentiment Analysis Influencing Word Embedding Using SVM and KNN," *Cyber Intell. Inf. Retr.*, 2021, doi: 10.1007/978-981-16-4284-5_18.
- [12] M. A. A. Bimbe, J. Marzal, and U. Khaira, "Comparison of K-Nearest Neighbor and Support Vector Machine Methods in Sentiment Analysis of Offline Courses," *Internet Things Artif. Intell. J.*, 2025, doi: 10.31763/iota.v5i1.898.