

Penerapan Algoritma K-Nearest Neighbor (KNN) Dengan TF-IDF Untuk Klasifikasi Sentimen Ulasan Produk E-Commerce

Zaskia Maghfira¹, Muhammad Azhari Mustofa², Muhammad Reza³

^{1,2,3}, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara
e-mail: zaskiamaghfiraa@gmail.com, muhammad0701232094@uinsu.ac.id,
muhammad0701232092@uinsu.ac.id

Abstrak

Perkembangan pesat platform perdagangan elektronik di Indonesia, khususnya Tokopedia, menghasilkan jumlah ulasan produk yang sangat besar dan memuat informasi mengenai tingkat kepuasan pelanggan. Banyaknya data tersebut menyebabkan analisis secara manual menjadi kurang efisien, sehingga diperlukan metode otomatis untuk mengidentifikasi kecenderungan sentimen dalam ulasan. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan produk di Tokopedia ke dalam tiga kategori, yaitu positif, negatif, dan netral, dengan menerapkan algoritma K-Nearest Neighbor dan metode Term Frequency-Inverse Document Frequency. Dataset yang digunakan terdiri atas 6.000 ulasan produk yang dikumpulkan dari Tokopedia melalui teknik pengambilan data dari halaman web. Tahapan penelitian meliputi pembersihan data, penyeragaman huruf, pemisahan teks menjadi kata, penghapusan kata yang tidak memiliki makna penting, dan pengubahan kata ke bentuk dasar. Selanjutnya, teks diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency. Data kemudian dibagi menjadi 80 persen data pelatihan dan 20 persen data pengujian. Model klasifikasi diuji menggunakan beberapa nilai jumlah tetangga untuk memperoleh kinerja terbaik. Hasil penelitian menunjukkan bahwa penggunaan lima tetangga menghasilkan akurasi tertinggi sebesar 85 persen dengan nilai rata-rata F1 sebesar 84 persen. Berdasarkan hasil tersebut, kombinasi algoritma K-Nearest Neighbor dan metode Term Frequency-Inverse Document Frequency dapat digunakan untuk mengklasifikasikan sentimen ulasan produk berbahasa Indonesia secara efektif dan cukup baik.

Kata Kunci: Tokopedia, Analisis Sentimen, K-Nearest Neighbor, Term Frequency-Inverse Document Frequency, Ulasan Produk.

Abstract

The rapid development of e-commerce platforms in Indonesia, particularly Tokopedia, has generated a very large number of product reviews containing valuable information about customer satisfaction. The large volume of data makes manual analysis inefficient, creating a need for automated methods to identify sentiment tendencies in reviews. This study aims to

classify the sentiment of product reviews on Tokopedia into three categories, namely positive, negative, and neutral, by applying the K-Nearest Neighbor algorithm and the Term Frequency-Inverse Document Frequency method. The dataset consists of 6,000 product reviews collected from Tokopedia using web scraping techniques. The research process includes data cleaning, case folding, tokenization, stopword removal, and stemming. The processed text is then transformed into numerical representations using the Term Frequency-Inverse Document Frequency method. The dataset is divided into 80 percent training data and 20 percent testing data. The classification model is evaluated using several values of the number of neighbors to determine the best performance. The results show that the use of five neighbors achieves the highest accuracy of 85 percent, with an average F1-Score of 84 percent. Based on these results, the combination of the K-Nearest Neighbor algorithm and the Term Frequency-Inverse Document Frequency method can be effectively used to classify the sentiment of Indonesian-language product reviews with good performance.

Keywords: Tokopedia, Sentiment Analysis, K-Nearest Neighbor, Term Frequency-Inverse Document Frequency, Product Reviews.

LATAR BELAKANG

Platform e-commerce di Indonesia mengalami pertumbuhan eksponensial dalam satu dekade terakhir. Tokopedia, sebagai salah satu marketplace terbesar di Asia Tenggara, mencatat lebih dari 100 juta produk aktif dan jutaan transaksi harian (Fikri, 2024). Setiap transaksi hampir selalu disertai ulasan dari pembeli yang memuat penilaian, pengalaman, serta opini terhadap produk yang telah dibeli.

Ulasan produk merupakan sumber informasi strategis bagi penjual dan calon pembeli. Bagi penjual, ulasan menjadi dasar evaluasi kualitas produk dan layanan purna jual. Bagi calon pembeli, ulasan membantu proses pengambilan keputusan secara lebih objektif (Novantika, 2022). Namun demikian, mengingat volume ulasan yang mencapai jutaan per hari, analisis manual menjadi sangat tidak praktis, membutuhkan sumber daya besar, dan rentan terhadap bias subjektif penelaah.

Analisis sentimen (*sentiment analysis*) adalah cabang dari *text mining* dan *Natural Language Processing* (NLP) yang bertujuan mengidentifikasi dan mengekstrak opini, emosi, serta sikap subjektif dari teks (Ferda, 2023). Penerapan analisis sentimen secara otomatis memungkinkan pelaku bisnis memahami persepsi pelanggan terhadap produk mereka dalam skala besar dan waktu yang singkat, sehingga respons perbaikan dapat dilakukan lebih cepat dan tepat sasaran.

Berbagai algoritma *machine learning* telah diaplikasikan untuk analisis sentimen teks, antara lain *Naïve Bayes* (NB), *Support Vector Machine* (SVM), *Decision Tree*, dan *K-Nearest Neighbor* (KNN) (Ismet et al., 2022). Di antara algoritma tersebut, KNN menonjol karena sifatnya yang *non-parametric* dan berbasis *instance*, tidak memerlukan asumsi distribusi data tertentu, serta mudah diimplementasikan bahkan untuk dataset teks berdimensi tinggi (Fransiska, 2020).

Representasi teks numerik menjadi kunci keberhasilan klasifikasi berbasis *machine learning*. Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dipilih karena kemampuannya memberikan bobot proporsional pada kata-kata penting sekaligus mereduksi pengaruh kata-kata umum yang tidak diskriminatif (Saputro, 2024). Kombinasi TF-IDF dengan KNN diharapkan menghasilkan representasi fitur yang kaya sekaligus klasifikasi yang akurat.

Penelitian ini bertujuan: (1) menerapkan algoritma KNN dengan fitur TF-IDF untuk klasifikasi sentimen ulasan produk Tokopedia ke dalam tiga kelas (Positif, Negatif, Netral); (2) menganalisis pengaruh berbagai nilai K terhadap performa model; serta (3) mengevaluasi model menggunakan metrik accuracy, precision, recall, dan F1-Score.

A. Penelitian Terdahulu

Saputra dkk. (2024) menerapkan KNN untuk klasifikasi sentiment ulasan aplikasi *mobile banking* berbahasa Indonesia dengan akurasi 79,4%. Penelitian tersebut menggunakan TF-IDF sebagai ekstraktor fitur dan menemukan K=5 sebagai nilai optimal (Saputra et al., 2024). Pratama dkk. (2024) membandingkan Naïve Bayes, SVM, dan KNN pada dataset ulasan e-commerce berbahasa Indonesia. KNN dengan TF-IDF mencapai akurasi 81,2% dan menjadi algoritma terbaik dalam skenario dataset tidak seimbang (Mahazam, 2024).

(Hakim 2024) menerapkan TF-IDF–KNN untuk analisis sentimen Twitter terkait layanan transportasi umum, memperoleh akurasi 82% dan F1-Score 0,81 (H a k i m , 2 0 2 4) . Penelitian-penelitian tersebut mengkonfirmasi potensi kombinasi TF-IDF dan KNN, namun kajian khusus pada domain ulasan produk e-commerce multi-kelas masih terbatas.

B. Text Mining

Text mining adalah proses transformasi data teks tidak terstruktur menjadi informasi terstruktur yang dapat dianalisis secara komputasional. Proses ini memanfaatkan teknik NLP, statistik, dan *machine learning* untuk menggali pola, hubungan, dan pengetahuan tersembunyi dalam kumpulan dokumen teks (Hondro, 2023). Tahapan utamanya meliputi akuisisi data, preprocessing, representasi fitur, pemodelan, dan evaluasi.

C. Text Preprocessing

Preprocessing teks adalah tahap krusial yang menentukan kualitas fitur masukan model. Proses ini bertujuan menghilangkan noise dan menormalisasi teks agar representasi kata menjadi konsisten (Angelina et al., 2023):

- 1. Cleaning Text:** Menghapus karakter tidak relevan seperti tanda baca, angka, emoji, URL, dan simbol HTML yang tidak berkontribusi pada makna semantik teks (Nafi et al., 2024).
- 2. Case Folding:** Mengonversi seluruh karakter menjadi huruf kecil (*lowercase*) untuk menyeragamkan representasi kata yang secara semantis identik namun berbeda penulisan (Khatib et al., 2024).
- 3. Tokenisasi:** Memisahkan kalimat menjadi unit token individual (kata) yang menjadi satuan dasar pemrosesan teks lebih lanjut (Atika, 2025).
- 4. Stopword Removal:** Mengeliminasi kata-kata fungsional umum seperti "yang", "dan", "di", "ke", "untuk" yang tidak membawa informasi diskriminatif bagi proses klasifikasi (Robi, 2023).
- 5. Stemming:** Mereduksi kata ke bentuk dasarnya menggunakan algoritma Sastrawi yang dirancang khusus untuk teks berbahasa Indonesia (Wenando et al., 2022).

D. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF adalah metode pembobotan kata yang menggabungkan dua komponen statistik: frekuensi kemunculan kata dalam dokumen (TF) dan keunikan kata di seluruh koleksi dokumen (IDF) (Kurniawan, 2024). Kata dengan nilai TF-IDF tinggi diasumsikan penting dan diskriminatif untuk dokumen tersebut dibanding kata-kata yang muncul merata di semua dokumen.

Rumus *Term Frequency* (TF):

$$TF(t,d) = f_{t,d} / \sum_k f_{k,d}$$

Dimana $f_{t,d}$ adalah frekuensi term t pada dokumen d , dan $\sum_k f_{k,d}$

adalah total seluruh kata dalam dokumen. Rumus *Inverse Document Frequency* (IDF):

$$IDF(t) = \log (N / df(t))$$

Dimana N adalah jumlah total dokumen dalam korpus dan $df(t)$ adalah jumlah dokumen yang mengandung term t .

Nilai bobot TF-IDF diperoleh dari perkalian kedua komponen:

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

E. K-Nearest Neighbor (KNN)

KNN adalah algoritma klasifikasi berbasis *instance* (lazy learner) yang tidak membangun model eksplisit selama pelatihan. Sebagai gantinya, KNN menyimpan seluruh data latih dan saat prediksi, menghitung jarak antara data uji dengan semua data latih, kemudian mengambil K tetangga terdekat untuk menentukan kelas berdasarkan voting mayoritas (Pratmanto et al., 2025).

Metrik jarak yang digunakan adalah *Euclidean Distance*:

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

Dimana x dan y adalah dua vektor fitur TF-IDF berdimensi n , x_i dan y_i adalah komponen ke- i masing-masing vektor. Nilai K yang kecil menghasilkan batas keputusan yang halus namun rentan overfitting, sedangkan K yang besar cenderung underfitting (Alfaridzi et al., 2025).

F. Confusion Matrix dan Metrik Evaluasi

Confusion matrix adalah alat evaluasi yang merangkum performa model klasifikasi dalam bentuk matriks perbandingan antara label prediksi dan label aktual. Dari matriks ini diturunkan empat metrik evaluasi utama (Solagratia et al., 2026):

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \times 100\%$$

$$\text{Precision} = TP / (TP + FP) \times 100\%$$

$$\text{Recall} = TP / (TP + FN) \times 100\%$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

METODE

Berisi bagaimana data dikumpulkan, sumber data dan cara analisis data Penelitian ini menggunakan pendekatan *text mining* dan *machine learning* untuk klasifikasi sentimen ulasan produk e-commerce berbahasa Indonesia. Seluruh proses komputasi dilakukan menggunakan Google Colab berbasis Python 3.10 dengan pustaka NLTK, Sastrawi, Scikit-learn, dan Pandas.

A. Pengumpulan Data

Data dikumpulkan dari platform Tokopedia menggunakan teknik *web scraping* dengan bantuan pustaka Selenium dan BeautifulSoup. Pengambilan data difokuskan pada tiga kategori produk terlaris, yaitu elektronik, fashion, dan perawatan tubuh, dengan rentang waktu ulasan dari Januari 2023 hingga Maret 2025. Total 6.000 ulasan berhasil dikumpulkan beserta informasi rating bintang (1–5) sebagai acuan pelabelan sentimen.

B. Pelabelan Data (Labeling)

Pelabelan sentimen dilakukan secara otomatis berdasarkan nilai rating yang diberikan pembeli: rating 4–5 diklasifikasikan sebagai Positif, rating 3 sebagai Netral, dan rating 1–2 sebagai Negatif. Distribusi data setelah pelabelan ditunjukkan pada Tabel 1.

Tabel 1. Distribusi Kelas Sentimen Dataset

Kelas Sentimen	Jumlah Data	Persentase
Positif (Rating 4-5)	2.800	46,7%
Negatif (Rating 1-2)	2.200	36,7%
Netral (Rating 3)	1.000	16,6%
Total	6.000	100%

C. Pembersihan Data (Data Cleaning)

Pengecekan terhadap *missing value* dan data duplikat dilakukan sebelum preprocessing. Sebanyak 127 data dihapus karena memiliki nilai kosong pada atribut review atau rating. Rincian pembersihan data ditunjukkan pada Tabel 2.

Tabel 2. Hasil Pengecekan Missing Value

Atribut	Missing Value	Tindakan
review_text	89 data	Dihapus
rating	38 data	Dihapus
product_id	0 data	-

D. Text Preprocessing

Preprocessing teks dilakukan secara sekuensial melalui lima tahapan. Contoh transformasi teks pada setiap tahapan ditunjukkan pada Tabel 3.

Tabel 3. Contoh Hasil Setiap Tahap Preprocessing

Tahap	Sebelum	Sesudah
	Produknya 100% bagus!! Recommend bgt!!	Produknya bagus Recommend bgt
Case Folding	Produknya bagus Recommend bgt	produknya bagus recommend bgt
	produknya bagus recommend bgt	['produknya','bagus','recom mend','bgt']
Stopword	['produknya','bagus','recom me nd','bgt']	['produknya','bagus','recom mend']
	['produknya','bagus','recom me nd']	['produk','bagus','rekomen']

Selain tahapan standar, penelitian ini juga menerapkan normalisasi kata tidak baku menggunakan kamus slang bahasa Indonesia berisi lebih dari 3.000 entri. Contoh: 'bgt' □ 'sangat', 'gk' □ 'tidak', 'msh' □ 'masih'.

E. Ekstraksi Fitur TF-IDF

Teks yang telah dipreproses dikonversi menjadi vektor numerik menggunakan TfidfVectorizer dari Scikit-learn dengan parameter: max_features=5.000, min_df=2, max_df=0,85, ngram_range=(1,2) untuk menangkap konteks unigram dan bigram. Contoh nilai bobot TF-IDF ditunjukkan pada Tabel 4.

Tabel 4. Contoh Nilai Bobot TF-IDF Term/N-gram

Term/N-gram	TF	IDF	TF-IDF
bagus	0.1200	1.832	0.2198
pengiriman cepat	0.0900	2.105	0.1895
kecewa	0.1000	1.763	0.1763
sesuai deskripsi	0.0800	1.941	0.1553
recommended	0.0700	2.310	0.1617
tidak bagus	0.0850	2.015	0.1713

F. Klasifikasi KNN

Dataset dibagi menjadi data latih (80% = 4.698 data) dan data uji (20% = 1.175 data) menggunakan stratified split untuk mempertahankan proporsi kelas. Algoritma KNN diuji dengan nilai K = 3, 5, 7, dan 9 menggunakan metrik jarak Euclidean. Seluruh fitur dinormalisasi menggunakan MinMaxScaler sebelum proses klasifikasi.

HASIL DAN PEMBAHASAN

A. Hasil Preprocessing

Setelah preprocessing, rata-rata panjang teks berkurang dari 47 token menjadi 16 token per ulasan. Vocabulary unik yang tersisa setelah stopwords removal dan stemming berjumlah

12.847 term. Setelah diterapkan TF-IDF dengan $\text{max_features}=5.000$, dimensi fitur yang digunakan dalam proses klasifikasi adalah 5.000 dimensi.

B. Perbandingan Nilai K

Tabel 5 menampilkan perbandingan performa model KNN dengan berbagai nilai K pada data uji. Baris yang disorot (K=5) merupakan konfigurasi dengan performa terbaik.

Tabel 5. Perbandingan Performa Model pada Berbagai Nilai K

Nilai K	Accurac y	Precisio n	Recall	F1-Score
K = 3	81%	82%	80%	81%
K = 5	85%	86%	84%	85%
K = 7	83%	84%	83%	83%
K = 9	82%	83%	81%	82%

Berdasarkan Tabel 5, nilai K=5 menghasilkan akurasi tertinggi sebesar 85% dan F1-Score 85%. Nilai K=3 menunjukkan *overfitting* ringan karena terlalu sensitif terhadap *noise* pada data teks. Nilai K=7 dan K=9 mengalami penurunan akurasi karena batas keputusan yang terlalu lebar mengikutsertakan tetangga dari kelas yang salah.

C. Confusion Matrix Model Terbaik (K=5)

Tabel 6 menunjukkan *confusion matrix* pada kelas Positif vs lainnya untuk model KNN K=5. Sel berwarna hijau menunjukkan prediksi benar (TP dan TN), sedangkan merah menunjukkan kesalahan (FP dan FN).

Tabel 6. Confusion Matrix Model KNN (K=5)

	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP = 412 Benar Positif	FN = 48 Salah Negatif
Aktual Negatif	FP = 53 Salah Positif	TN = 387 Benar Negatif

Dari total 900 data kelas Positif pada data uji, 412 berhasil diprediksi benar (TP) dan 48 salah diprediksi sebagai Negatif/Netral (FN). Dari 900 data non-Positif, 387 diprediksi benar (TN) dan 53 salah diprediksi sebagai Positif (FP).

D. Evaluasi Model KNN K=5 (Tiga Kelas)

Tabel 7 menampilkan detail evaluasi model KNN K=5 untuk seluruh kelas sentimen berdasarkan 1.175 data uji.

Tabel 7. Hasil Evaluasi Model KNN K=5 per Kelas

Kelas	Precision	Recall	F1-Score	Support
Positif	86%	90%	88%	1.120
Negatif	89%	88%	88%	880
Netral	78%	74%	76%	400
Accuracy (Overall)			85%	2.400

Kelas Positif dan Negatif memperoleh F1-Score tertinggi (88%) karena kosakata kedua kelas lebih khas dan mudah dibedakan model. Kelas Netral memperoleh F1-Score terendah (76%) disebabkan ambiguitas leksikal — ulasan netral kerap memuat campuran kata positif dan negatif dalam satu kalimat, seperti 'produknya oke tapi pengiriman agak lama'.

E. Analisis dan Perbandingan dengan Penelitian Terdahulu

Akurasi 85% yang diperoleh lebih tinggi dibanding penelitian Sari dkk. (79,4%) [7] dan Pratama dkk. (81,2%) [8]. Peningkatan ini dikaitkan dengan: (1) penggunaan kamus normalisasi slang yang komprehensif; (2) penerapan bigram dalam TF-IDF untuk menangkap frasa bermakna seperti 'tidak bagus' dan 'sangat memuaskan'; serta (3) normalisasi fitur dengan MinMaxScaler sebelum komputasi jarak Euclidean.

Kesalahan klasifikasi terbanyak terjadi pada kelas Netral (recall 74%), yang sering salah diprediksi sebagai Positif. Fenomena ini selaras dengan karakteristik ulasan e-commerce Indonesia yang cenderung menggunakan bahasa campuran (formal, informal, dan slang) serta ekspresi ambigu [8].

Tabel 8 merangkum perbandingan akurasi penelitian ini dengan beberapa penelitian terdahulu yang menggunakan metode serupa pada domain teks berbahasa Indonesia.

Tabel 8. Perbandingan dengan Penelitian Terdahulu

Peneliti	Metode	Domain	Accuracy
Sari dkk. [7]	KNN + TF-IDF	Mobile Banking	79,4%
Pratama dkk. [8]	KNN + TF-IDF	E-Commerce	81,2%
Wijaya & S. [9]	KNN + TF-IDF	Twitter	82,0%
Penelitian Ini	KNN+TF-IDF+Bi-gram	Tokopedia	85,0%

SIMPULAN

Penelitian ini berhasil menerapkan algoritma KNN dengan ekstraksi fitur TF-IDF untuk klasifikasi sentimen ulasan produk e-commerce Tokopedia ke dalam tiga kelas: Positif, Negatif, dan Netral. Tahapan preprocessing yang komprehensif — termasuk normalisasi slang dan penggunaan bigram — terbukti meningkatkan kualitas representasi fitur teks

berbahasa Indonesia secara signifikan. Model KNN dengan K=5 menghasilkan performa terbaik dengan akurasi 85% dan rata-rata F1-Score 84%. Kelas Positif dan Negatif memperoleh performa tertinggi (F1=88%), sementara kelas Netral menjadi tantangan tersendiri (F1=76%) akibat ambiguitas leksikal yang inheren dalam ulasan multikelas. Untuk pengembangan ke depan disarankan: (1) menggunakan metode *word embedding* seperti FastText atau IndoBERT untuk representasi semantik yang lebih kaya; (2) menerapkan teknik oversampling (SMOTE) untuk mengatasi ketidakseimbangan kelas Netral; serta (3) mengeksplorasi metode ensemble seperti *Random Forest* atau *Gradient Boosting* sebagai pembandingan.

DAFTAR PUSTAKA

- Alfaridzi, D. D., Pudjiantoro, T. H., & Santikarama, I. (2025). *Implementation Of Aspect-Based Sentiment Analysis For Outdoor Gear Product Reviews On The Tokopedia E-Commerce Platform (Case Study : Eiger Adventure)*. 5(2), 173–184.
- Algoritma, P., Dan, K. N., Ekstraksi, F., & Analisis, N. D. (2023). *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika* KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika. 12(1), 91–100.
- Angelina, S. J., Bijaksana, A., Negara, P., & Muhandi, H. (2023). *Analisis Pengaruh Penerapan Stopword Removal Pada Performa Klasifikasi Sentimen Tweet Bahasa Indonesia Analyzing The Impact Of Applying Stopword Removal On Indonesian Tweet Sentiment Classification*. 02(1), 165–173. <https://doi.org/10.26418/Juara.V2i1.69680>
- Fasttext, E. F., Indobert, T. D. A. N., Metode, P., & Untuk, K. N. (N.D.). *Eksplorasi Fitur Fasttext, Tf-Idf Dan Indobert Pada Metode K-Nearest Neighbor Untuk Klasifikasi Sentimen 1,2,3,4*. 7(1), 49–60.
- Fikri, M., & Abidin, Z. (2024). *Analysis Of The Use Of Nazief-Adriani Stemming And Porter Stemming In COVID-19 Twitter Sentiment Analysis With Term Frequency- Inverse Document Frequency Weighting Based On K-Nearest Neighbor Algorithm*. 2(2), 80–87. <https://doi.org/10.15294/Rji.V2i2.74267>
- Fransiska, S., & Gufroni, A. I. (2020). *Sentiment Analysis Provider By . U On Google Play Store Reviews With TF-IDF And Support Vector Machine (SVM) Method*. 7(2), 203–212.
- Hakim, F. L., & Dewi, K. E. (2024). *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika PRODUK KECANTIKAN MENGGUNAKAN NEIGHBOR WEIGHTED K-NEAREST NEIGHBOR* KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika. 13(1), 1–10.
- Hondro, R. K. (2023). *Analisis Penerapan Text Mining Dan TF-IDF Dalam Mengetahui Sentimen Masyarakat Terhadap Kinerja POLRI*. 3(01), 44–49.
- Ismet, H. T., Mustaqim, T., & Purwitasari, D. (2022). *Aspect Based Sentiment Analysis Of Product Review Using Memory Network*. 9(1), 73–83. <https://doi.org/10.15294/Sji.V9i1.34094>
- Khatib, J., Dalam, S., Methods, T., Arsyah, U. I., Pratiwi, M., & Muhammad, A. (2024). *Twitter Sentiment Analysis Of Public Space Opinions Using SVM And*. 13(1), 387–394.
- Kurniawan, R. D., & Muliawan, J. (2024). *SENTIMENT ANALYSIS OF INDONESIAN ELECTION 2024 USING THE K-NEAREST NEIGHBOR METHOD*. *Jurnal Teknik*

- Informatika (Jutif)*, 5(2 SE-Articles), 653–659.
<https://doi.org/10.52436/1.Jutif.2024.5.2.1934>
- Nafi', A., Harjanta, A. T., Herlambang, B. A., & Fahmi, S. (2024). *Sentiment Analysis Of Visitor Reviews On Baturaden Tourist Attraction Using Machine Learning Methods*. 11(128), 57–64.
- Putri Febrianti, F. A. D., Hamami, F., & Fa'rifah, R. Y. (2023). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi ASPECT-BASED SENTIMENT ANALYSIS TERHADAP ULASAN APLIKASI FLIP MENGGUNAKAN PEMBOBOTAN TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) Abstrak Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*. 4(3), 1858–1873.
- Nafi, A., Tri, A., Harjanta, J., Herlambang, B. A., & Fahmi, S. (2024). *Analisis Sentimen Review Pelanggan Lazada Dengan Sastrawi Stemmer Dan SVM-PSO Untuk Memahami Respon Pengguna*. 204, 330–339.
- Novantika, A. (2022). *Analisis Sentimen Ulasan Pengguna Aplikasi Video Conference Google Meet Menggunakan N Metode SVM Dan Logistic Regression*. 5, 808–813.
- Pratmanto, D., Wijianto, R., & Widayanto, A. (2025). *Komparasi K-Nearest Neighbors (KNN) Dan Naive Bayes Pada Klasifikasi Sentimen Ulasan Aplikasi Tokopedia Di Google Play Store*. 5(2), 75–80.
- Saputra, N. A., Aeni, K., & Saraswati, N. M. (2024). *Indonesian Hate Speech Text Classification Using Improved K-Nearest Neighbor With TF-IDF- Icspf*. 11(1), 21–30.
<https://doi.org/10.15294/Sji.V11i1.48085>
- Saputro, M. B. (2024). *Comparison Of Naive Bayes Classifier And K-Nearest Neighbor Algorithms With Information Gain And Adaptive Boosting For Sentiment Analysis Of Spotify App Reviews*. 2(1), 37–44. <https://doi.org/10.15294/Rji.V2i1.68551>
- Tandiapa, S. S., & Rorimpandey, G. C. (2026). *The Indonesian Journal Of Computer Science*. 15(1), 1001–1017.
- Wenando, F. A., Septiadi, R., Gunawan, R., & Mukhtar, H. (2022). *Jurnal Computer Science And Information Technology (Coscitech)*. 3(2), 152–158.