

---

---

## PENERAPAN FEATURE SELECTION UNTUK MENINGKATKAN KINERJA METODE NAIVE BAYES DALAM ANALISIS SENTIMEN KONSUMEN

**Imelda Juniarti Layuk<sup>1</sup>, Muhammad Arafah<sup>2</sup>, Andi Maulidinnawati Abdul Kadir  
Parewe<sup>3</sup>, Ilham<sup>4</sup>, Pasnur<sup>5</sup>, Muhajirin<sup>6</sup>, Tatik Maslihatin<sup>7</sup>**

<sup>1,2,3,4,5,6,7</sup>Universitas Teknologi Akba Makassar

Jl. Perintis Kemerdekaan, Tamalanrea, Kec. Tamalanrea, Kota Makassar, Sulawesi Selatan  
amaulidinnawati@akba.ac.id\*

### **ABSTRACT**

*The rapid growth of social commerce platforms such as TikTok Shop has significantly increased consumer interactions, generating a vast number of opinions and product reviews in real time. These reviews serve as valuable sources of information for companies to better understand consumer perceptions of their products. This study aims to analyze consumer sentiment toward Wardah Cushion products on the TikTok Shop platform by employing the Naive Bayes method and applying Feature Selection techniques. The analysis was conducted under two scenarios: without feature selection and with feature selection using the Chi-Square method. The data preprocessing stage involved cleaning, case folding, stopword removal, stemming, and tokenization to prepare the text before classification. Model performance was evaluated using accuracy, precision, recall, and f1-score metrics. The experimental results show that the application of feature selection improved model accuracy from 88.84% to 92.15%. Based on these findings, it can be concluded that selecting relevant features has a positive impact on the performance of the Naive Bayes model in classifying consumer sentiment more accurately. This research contributes to the utilization of text mining and machine learning in sentiment analysis for social media-based e-commerce platforms. For future work, it is recommended to employ larger datasets and compare different classification algorithms to obtain more comprehensive results.*

**Keywords:** Sentiment Analysis, Naive Bayes, Feature Selection, Chi-Square, TikTok Shop

### **ABSTRAK**

E-commerce berbasis media sosial, seperti TikTok Shop, menghasilkan opini konsumen yang penting sebagai sumber informasi bagi perusahaan. Penelitian ini menganalisis sentimen konsumen terhadap produk Wardah Cushion di TikTok Shop menggunakan metode Naive Bayes dengan dan tanpa penerapan feature selection Chi-Square. Data ulasan diproses melalui tahapan cleaning, case folding, stopword removal, stemming, dan tokenization sebelum diklasifikasikan. Kinerja model dievaluasi menggunakan akurasi, precision, recall, dan f1-score. Hasil menunjukkan bahwa feature selection meningkatkan akurasi model dari 88,84% menjadi 92,15%, sehingga terbukti efektif dalam meningkatkan performa Naive Bayes dengan mengurangi fitur yang tidak relevan. Temuan ini menegaskan pentingnya pemilihan fitur dalam analisis sentimen untuk menghasilkan klasifikasi yang lebih akurat dan efisien. Penelitian ini berkontribusi pada pengembangan text mining dan machine learning di e-commerce berbasis media sosial. Untuk penelitian selanjutnya, disarankan menggunakan dataset lebih besar dan

## PENDAHULUAN

Teknologi informasi telah menjadi bagian yang tidak terpisahkan dari kehidupan manusia modern. Di era digital, perkembangan teknologi bukan hanya memengaruhi sektor bisnis dan pemerintahan, tetapi juga kehidupan masyarakat sehari-hari. Penggunaan internet yang semakin meluas telah menciptakan budaya baru di mana hampir seluruh lapisan masyarakat memanfaatkan teknologi digital untuk komunikasi, hiburan, pendidikan, hingga kegiatan ekonomi. Kondisi ini menunjukkan bahwa transformasi digital telah merambah seluruh aspek kehidupan, sehingga kebutuhan untuk memahami dan mengelola informasi yang tersedia dalam jumlah besar menjadi semakin penting.

Salah satu bentuk nyata dari penetrasi teknologi informasi adalah munculnya media sosial sebagai ruang interaksi digital yang populer. Media sosial, yang awalnya hanya difungsikan sebagai sarana berbagi momen pribadi, kini telah berkembang menjadi platform multifungsi. Berbagai aplikasi media sosial tidak hanya digunakan untuk berkomunikasi, tetapi juga untuk berbisnis, mencari informasi, hingga membentuk opini publik. TikTok, sebagai salah satu aplikasi media sosial yang berkembang pesat, merepresentasikan fenomena tersebut. Popularitas TikTok semakin meningkat dari tahun ke tahun dengan jumlah pengguna aktif yang terus bertambah (Widagdo et al., 2020). Aplikasi ini awalnya dikenal dengan konten hiburan berupa video pendek dengan tambahan musik dan filter menarik, tetapi dalam perkembangannya TikTok memperluas fungsinya dengan menyediakan fitur baru yang mendukung aktivitas bisnis.

Fitur inovatif tersebut dikenal sebagai TikTok Shop, yaitu sebuah layanan e-commerce yang terintegrasi langsung di dalam aplikasi TikTok. TikTok Shop memungkinkan pengguna untuk melakukan transaksi belanja tanpa harus berpindah ke aplikasi lain. Hal ini merupakan inovasi signifikan dalam dunia bisnis online karena memadukan interaksi sosial dengan transaksi komersial. Dengan adanya TikTok Shop, pengguna dapat membeli produk langsung dari video yang mereka tonton, sehingga menciptakan pengalaman belanja yang lebih praktis dan interaktif (Maulana et al., 2024). Tidak hanya itu, fitur ini juga memberikan peluang besar bagi pelaku usaha, baik skala kecil maupun besar, untuk memasarkan produknya secara langsung kepada konsumen potensial dengan cara yang lebih personal dan menarik.

Fenomena TikTok Shop memperlihatkan bagaimana teknologi informasi mendorong terciptanya model bisnis baru yang memadukan hiburan dan perdagangan. Namun, kehadiran fitur ini juga menimbulkan beragam respon dari masyarakat. Sebagian pengguna memberikan opini positif dengan menilai bahwa TikTok Shop memudahkan proses belanja, menyediakan beragam pilihan produk, serta menghadirkan harga yang kompetitif. Di sisi lain, ada pula opini negatif yang mempertanyakan kualitas produk, keamanan transaksi, dan efektivitas layanan. Selain itu, tidak sedikit pula opini netral yang hanya membandingkan TikTok Shop dengan e-marketplace lain seperti Shopee, Tokopedia, atau Lazada. Keragaman opini tersebut menunjukkan kompleksitas pandangan masyarakat terhadap TikTok Shop sebagai fitur belanja online baru (Indriyani et al., 2023).

Opini publik yang terekam di media sosial sangat berharga bagi perusahaan. Ulasan konsumen dapat menjadi cerminan kualitas produk dan layanan yang ditawarkan. Bagi perusahaan kosmetik seperti Wardah, opini yang berkembang di TikTok Shop menjadi salah satu tolok ukur penting untuk memahami bagaimana konsumen menilai produk mereka. Hal ini karena industri kosmetik sangat erat kaitannya dengan persepsi konsumen terhadap kualitas, keamanan, serta kecocokan produk. Oleh karena itu, analisis terhadap opini masyarakat bukan hanya relevan, tetapi juga strategis bagi perusahaan yang ingin meningkatkan daya saing produknya di pasar digital.

Beragamnya opini konsumen di media sosial menimbulkan kebutuhan untuk mengolah data dalam jumlah besar secara sistematis. Di sinilah analisis sentimen memainkan peran penting. Analisis sentimen merupakan pendekatan analitis yang digunakan untuk mengidentifikasi kecenderungan positif, negatif, atau netral dari suatu teks. Teknik ini, yang juga dikenal dengan istilah opinion mining, memanfaatkan metode text mining dan machine learning untuk menggali informasi dari data tidak terstruktur. Dibiidang e-commerce, analisis sentimen digunakan untuk menilai persepsi konsumen terhadap produk atau layanan tertentu.

Urgensi analisis sentimen semakin besar di era big data, di mana ulasan konsumen tersedia dalam jumlah yang sangat masif. Tanpa metode analitis yang tepat, perusahaan akan kesulitan untuk memahami pola dan tren yang muncul dari opini konsumen. Dengan analisis sentimen, data ulasan yang semula bersifat kualitatif dapat dikonversi menjadi informasi

kuantitatif yang berguna untuk pengambilan keputusan. Misalnya, jika mayoritas ulasan menunjukkan sentimen positif, perusahaan dapat menggunakannya sebagai dasar untuk memperkuat strategi pemasaran. Sebaliknya, jika ditemukan banyak sentimen negatif, perusahaan dapat segera melakukan evaluasi terhadap produk atau layanannya.

Selain untuk perusahaan, analisis sentimen juga bermanfaat bagi konsumen. Ulasan yang dianalisis secara sistematis dapat membantu konsumen lain dalam menentukan pilihan produk. Produk Wardah Cushion di TikTok Shop, analisis sentimen dapat memberikan gambaran seberapa besar tingkat kepuasan konsumen terhadap produk tersebut, apakah lebih banyak yang puas atau justru tidak. Dengan demikian, analisis sentimen berfungsi sebagai sarana transparansi informasi yang dapat meningkatkan kepercayaan antara produsen, konsumen, dan platform e-commerce itu sendiri.

Penelitian sebelumnya telah menunjukkan bahwa analisis sentimen menjadi salah satu metode yang efektif dalam memahami opini masyarakat. Misalnya, penelitian yang dilakukan oleh Fatmawati et al. (2024) menjelaskan bahwa analisis sentimen dapat mengidentifikasi subjektivitas dari sebuah pendapat, ulasan, atau tweet sehingga dapat digunakan sebagai dasar bagi organisasi dalam merancang strategi yang lebih tepat. Analisis sentimen juga terbukti bermanfaat dalam berbagai bidang lain, seperti politik untuk memantau opini publik, kesehatan untuk memahami pengalaman pasien, hingga sistem rekomendasi dalam aplikasi digital. Fakta ini menunjukkan bahwa analisis sentimen memiliki cakupan yang luas dan dapat diadaptasi sesuai dengan kebutuhan penelitian maupun bisnis.

Pada Jejaring sosial TikTok Shop di Indonesia, penelitian terkait analisis sentimen masih relatif baru. Sebagian besar penelitian sebelumnya lebih banyak berfokus pada platform e-commerce konvensional seperti Shopee atau Tokopedia. Padahal, TikTok Shop memiliki karakteristik yang unik karena menggabungkan interaksi sosial dengan transaksi belanja. Oleh karena itu, penelitian ini memfokuskan pada eksplorasi opini konsumen terhadap produk kosmetik, khususnya Wardah Cushion, di TikTok Shop. Dengan menggunakan teknik analisis sentimen, diharapkan dapat terungkap pola, preferensi, dan persepsi konsumen secara lebih jelas.

## **METODE PENELITIAN**

Metode Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional yang berfokus pada analisis sentimen konsumen terhadap produk Wardah Cushion di platform e-commerce berbasis media sosial, yaitu TikTok Shop. Pemilihan pendekatan ini didasarkan pada kebutuhan untuk mengukur kinerja algoritma klasifikasi dalam mengolah data teks secara sistematis dan terukur. Data penelitian berupa komentar konsumen yang diambil dari ulasan produk dengan teknik web scraping. Tahap ini penting untuk memperoleh data asli secara langsung dari sumbernya, sehingga representasi opini konsumen dapat tergambar dengan baik.

Setelah data terkumpul, dilakukan proses preprocessing teks untuk menyiapkan data agar lebih terstruktur sebelum dianalisis lebih lanjut. Tahapan preprocessing meliputi beberapa langkah utama. Pertama, cleaning, yaitu pembersihan data dari karakter atau simbol yang tidak relevan. Kedua, case folding, yaitu mengubah seluruh teks menjadi huruf kecil agar konsistensi data terjaga. Ketiga, stopword removal, yakni menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis, seperti kata hubung atau kata ganti. Keempat, stemming, yaitu mengembalikan kata pada bentuk dasarnya. Terakhir, tokenization, yakni memecah kalimat menjadi unit kata yang lebih kecil. Tahapan ini memastikan bahwa data teks siap diolah dan dianalisis secara lebih efektif.

Setelah preprocessing, data teks yang telah bersih diekstraksi menggunakan metode Bag of Words (BoW). Metode ini dipilih karena sederhana dan efektif dalam merepresentasikan frekuensi kemunculan kata pada dokumen. Untuk meningkatkan kualitas representasi fitur, penelitian ini membandingkan dua skenario. Skenario pertama adalah klasifikasi tanpa menggunakan feature selection, sehingga semua fitur dalam dataset dipertahankan. Skenario kedua menggunakan metode feature selection berbasis Chi-Square, yang berfungsi untuk memilih kata-kata paling relevan dalam membedakan sentimen positif dan negatif. Dengan adanya feature selection, jumlah fitur dapat dikurangi tanpa menghilangkan informasi penting, sehingga kinerja model diharapkan lebih optimal.

Proses klasifikasi dilakukan menggunakan algoritma Naïve Bayes. Pemilihan algoritma ini didasarkan pada kesederhanaan, efisiensi, serta efektivitasnya dalam menangani data teks yang besar dan bersifat probabilistik. Algoritma Naïve Bayes bekerja dengan

Imelda Juniarti Layuk,dkk

menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan distribusi fitur yang ada. Dalam konteks penelitian ini, algoritma digunakan untuk mengklasifikasikan komentar konsumen ke dalam kategori sentimen positif atau negatif.

Evaluasi kinerja model dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, precision, recall, dan f1-score. Akurasi digunakan untuk mengukur tingkat keseluruhan prediksi yang benar, sedangkan precision menilai ketepatan model dalam memprediksi komentar yang relevan. Recall mengukur kemampuan model dalam menemukan seluruh data relevan, dan f1-score menjadi ukuran keseimbangan antara precision dan recall. Keempat metrik ini dipilih agar evaluasi lebih komprehensif dan tidak hanya berfokus pada satu aspek kinerja saja.

## **HASIL DAN PEMBAHASAN**

### **Hasil**

Hasil penelitian menggambarkan temuan utama dari penelitian melalui Langkah sebagai berikut:

#### **1. Metode Klasifikasi**

metode klasifikasi Naive Bayes untuk melakukan analisis sentimen dengan tujuan mengelompokkan ulasan atau komentar konsumen terhadap produk Wardah Cushion di TikTok Shop ke dalam tiga kategori sentimen: Positif, Negatif, dan Netral. Setiap komentar diubah menjadi bentuk representasi numerik (bag-of-words atau TF-IDF) yang kemudian digunakan untuk menghitung probabilitas setiap komentar termasuk ke masing-masing kelas sentimen.

#### **2. Data training dan data testing**

Dataset yang digunakan pada penelitian ini terdiri dari 1210 data ulasan, yang dibagi menjadi dua bagian yaitu data latih sebesar 80% dan data uji sebesar 20% menggunakan fungsi `train_test_split` dari pustaka Scikit-learn. Pembagian ini bertujuan untuk mengukur performa model terhadap data yang belum pernah dilihat sebelumnya, guna menghindari overfitting dan mendapatkan evaluasi yang lebih objektif.

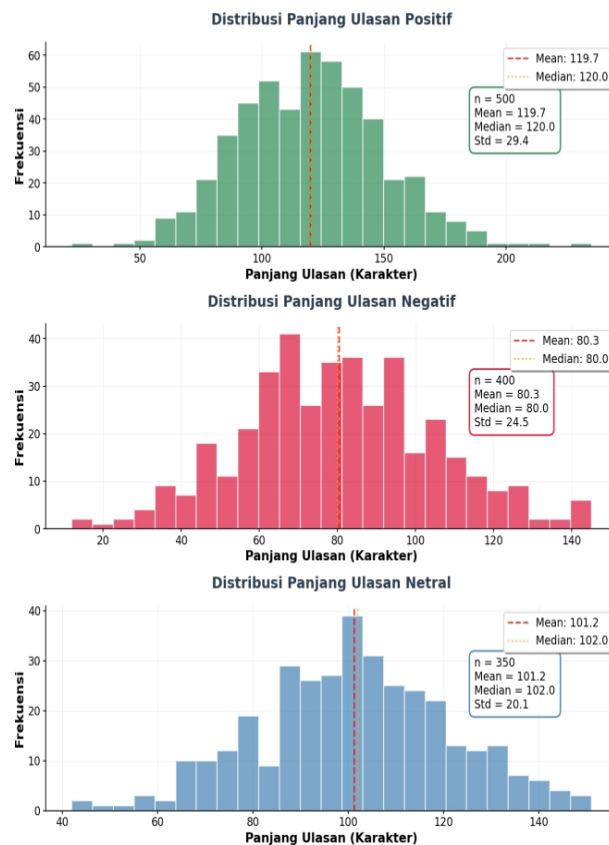
#### **3. Proses Scraper/ web scraping**

Web scraping merupakan teknik pengambilan data secara otomatis dari website. Teknik ini digunakan karena data yang tersedia pada halaman produk TikTok Shop yang kini terintegrasi dengan Tokopedia bersifat dinamis, dalam jumlah besar, dan tidak

disediakan secara langsung dalam bentuk dataset publik. Dengan web scraping, peneliti dapat memperoleh informasi aktual seperti nama produk, harga, deskripsi, maupun ulasan pelanggan secara efisien dan konsisten untuk kebutuhan analisis khusus Tokopedia. Penelitian ini, data dikumpulkan dari TikTok Shop dengan mengambil ulasan atau komentar konsumen terkait produk Wardah Cushion. Proses scraping dilakukan menggunakan bahasa pemrograman Python dan tools seperti BeautifulSoup atau Selenium.

#### 4. Pelabelan data/ history teks ulasan

Gambar dibawah ini menampilkan hasil pelabelan data atau dapat juga disebut dengan histogram teks ulasan yang merepresentasikan (review).



Gambar 1. Analisis distribusi Panjang dan ulasan berdasarkan sentimen

#### 5. Cleaning

Pada tabel 1, ulasan positif cenderung lebih panjang dan detail dengan rata-rata 119,7 karakter, sedangkan ulasan negatif lebih singkat dan langsung ke inti masalah dengan rata-rata 80,3 karakter. Ulasan netral berada di posisi tengah dengan rata-rata 101,2 karakter dan konsistensi yang tinggi. Pola ini menunjukkan bahwa panjang ulasan

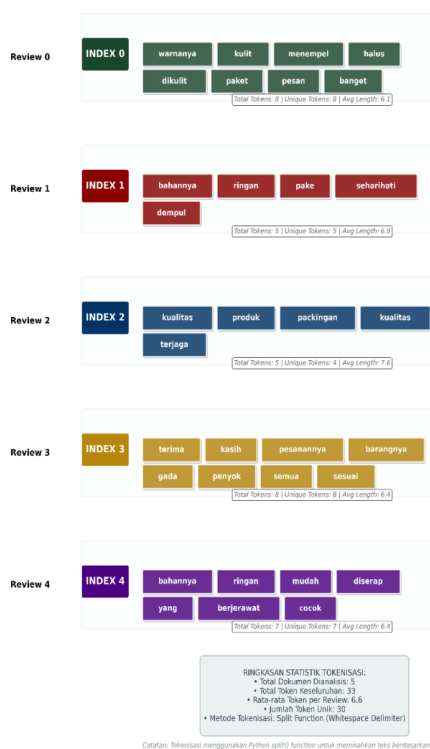
berkorelasi dengan jenis sentimen, di mana ulasan positif lebih ekspresif, negatif lebih ringkas, dan netral lebih stabil serta informatif.

Tabel 1. Rata-rata ulasan berdasarkan sentimen

Sentimen	Rata-rata	Median	Standar Deviasi	Karakter utama
Positif	119,7 karakter	120	29,4	Panjang konsisten, detail, ekspresif
Negatif	80,3 karakter	80	24,5	Singkat, jelas, langsung ke inti masalah
Netral	101,2 karakter	102	20,1	Moderat, konsisten, deskriptif, informatif, stabil

## 6. Tokennization

Tokenization adalah proses memecah teks menjadi unit -unit kecil yang disebut token, umumnya berupa kata.



Gambar 2. Hasil Tokenization

Gambar 2 yaitu proses tokenisasi menggunakan split function berhasil memecah 5 ulasan produk menjadi 33 token dengan rata-rata 6,6 token per ulasan dan 30 token unik, menunjukkan keragaman kosakata yang baik. Distribusi panjang token berada pada kisaran 6,1–7,6 karakter, sesuai dengan karakteristik bahasa Indonesia. Ulasan indeks 2 dan 4 memiliki token unik lebih banyak, sedangkan ulasan lain menunjukkan pengulangan kata yang wajar dalam konteks e-

commerce. Hasil ini membuktikan bahwa metode split function efektif untuk tahap awal preprocessing sebelum case folding, stopwords removal, dan filtering.

### 7. Case Folding

Tabel 2. proses case folding

Kalimat asli	Setelah Case Folding
"Produk Ini Sangat BAGUS!" ..... Pengiriman LAMBAT dan Tidak Direkomendasikan!!!"	"produk ini sangat bagus" ..... "Pengiriman lambat dan tidak direkomendasikan"

### 8. Stopword

	text	label	hapus
0	wamanya sesuai kulit menempel halus dikulit p...	positif	wamanya sesuai kulit menempel halus dikulit p...
1	bahannya ringan cocok untuk di pake seharihati...	positif	bahannya ringan cocok pake seharihati dempul
2	kualitas produk bagus packingan aman sehingga ...	positif	kualitas produk bagus packingan aman kualitas ...
3	terima kasih wardah pesanannya sesuai barangny...	positif	terima kasih wardah pesanannya sesuai barangny...
4	bahannya ringan mudah diserap cocok buat yang ...	positif	bahannya ringan mudah diserap cocok yang berje...
...	...	...	...
1205	mantap pokoknya cepet recommended banget	positif	mantap pokoknya cepet recommended banget
1206	bagus hasilnya kelihatan natural	positif	bagus hasilnya natural
1207	games bngt dapet pouch lucu terima kasih warda...	positif	games bngt dapet pouch lucu terima kasih warda...
1208	pesanan sesuai dateng dengan cepat dan selamat	positif	pesanan sesuai dateng cepat selamat
1209	cocok bgtt dikulit ade akuuu dia beli lewat ak...	positif	cocok bgtt dikulit ade akuuu beli akun promo

Gambar 3. Hasil Stopword

### 9. Stemming

Stemming adalah proses mengubah kata menjadi bentuk dasarnya. Dalam bahasa Indonesia, proses stemming akan menghapus imbuhan seperti "me-", "di-", "-kan", "-an", dan lainnya

### 10. Model latihan naive bayes

## KESIMPULAN

Membuktikan bahwa metode Naïve Bayes efektif untuk analisis sentimen konsumen terhadap produk Wardah Cushion di TikTok Shop. Penerapan feature selection dengan Chi-Square meningkatkan akurasi model hingga 92,15% serta membuat proses klasifikasi lebih efisien. Meski terdapat kendala pada distribusi kelas negatif, secara keseluruhan kombinasi Naïve Bayes dan feature selection mampu menghasilkan model analisis sentimen yang lebih akurat, efisien, dan bermanfaat dalam memahami persepsi konsumen.

## DAFTAR PUSTAKA

- Fatmawati, F., Irawan, B., & Bahtiar, A. (2024). Analisis Sentimen Pengguna Aplikasi Shejek Berdasarkan Ulasan Di Google Play Store Menggunakan Metode Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 2976–2981. <https://doi.org/10.36040/jati.v8i3.9607>
- Friska Aditia Indriyani, Ahmad Fauzi, & Sutan Faisal. (2023). Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine. *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, 10(2), 176–184. <https://doi.org/10.37373/tekno.v10i2.419>
- Indriati, I.-, Rahayudi, B., & Dewi, C. (2021). Analisis Sentimen Mengenai Moda Raya Terpadu (MRT) Jakarta dengan Metode BM25 dan K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(2), 389394. <https://doi.org/10.25126/jtiik.2021824508>
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naïve Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384. <https://doi.org/10.57152/malcom.v4i2.1206>
- Purnamawati, A., Winarto, M. N., & Mailasari, M. (2023). Analisis Sentimen Aplikasi TikTok menggunakan Metode BM25 dan Improved K-NN Fitur Chi-Square. *Jurnal Komtika (Komputasi Dan Informatika)*, 7(1), 97–105. <https://doi.org/10.31603/komtika.v7i1.8938>
- Ramadhan, B. Z., Adam, R. I., & Maulana, I. (2022). Analisis Sentimen Ulasan pada Aplikasi E-Commerce dengan Menggunakan Algoritma Naïve Bayes. *Journal of Applied Informatics and Computing*, 6(2), 220–225. <https://doi.org/10.30871/jaic.v6i2.4725>
- Sanjaya, R., Tohidi, E., Wahyudi, E., & Kaslani, K. (2024). Analisis Sentimen Terhadap Berhentinya TikTok Shop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 507–514. <https://doi.org/10.36040/jati.v8i1.8443>
- Saputra, I., AJI PAMBUDI, R. S., DARONO, H. E., AMSURY, F., FAHDIA, M. R., RAMADHAN, B., & ARDIANSYAH, A. (2021). Analisis Sentimen Pengguna Marketplace Bukalapak dan Tokopedia di Twitter Menggunakan Machine Learning. *Faktor Exacta*, 13(4), 200. <https://doi.org/10.30998/faktorexacta.v13i4.7074>

- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *Journal of Intelligent System and Computation*, 1(1), 43–49. <https://doi.org/10.52985/insyst.v1i1.36>
- Widagdo, A. S., W.A, B. S., & Nasiri, A. (2020). Analisis Tingkat Kepopuleran E Commerce Di Indonesia Berdasarkan Sentimen Sosial Media Menggunakan Metode Naïve Bayes. *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, 6(1), 1–5. <https://doi.org/10.46808/informa.v6i1.159>