

Customer Sentiment Analysis Toward Msme Products On Tiktok Social Media Using Naive Bayes Algorithm

Muhammad Husni Rifqo ¹⁾; Elezah Syafitri ²⁾; Khairullah ³⁾; Ujang Juhardi ⁴⁾

^{1,2,3,4)} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu

Email: ¹⁾ mhrifqo@umb.ac.id

How to Cite :

Rifqo, M. H., Syafitri, E., Khairullah, Y., Juhardi, U. (2026). Customer Sentiment Analysis Toward MSME Products on Tiktok Social Media Using Naive Bayes Algorithm. Jurnal Media Computer Science, 5(4).

ARTICLE HISTORY

Received [31 Agustus 2026]

Revised [25 Oktober 2026]

Accepted [28 Oktober 2026]

KEYWORDS

Sentiment Analysis, TikTok, MSME,
Naive Bayes, Preprocessing, TF-IDF.

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



ABSTRAK

Pesatnya perkembangan media sosial TikTok menjadikannya sarana promosi digital yang efektif bagi Usaha Mikro, Kecil, dan Menengah (UMKM). —Banyaknya komentar yang ditinggalkan oleh pengguna menjadi sumber informasi penting untuk mengukur persepsi dan kepuasan pelanggan. Namun, karakteristik bahasa teks media sosial yang tidak terstruktur, informal, mengandung kata slang, dan emoji menyulitkan proses analisis secara manual. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap produk UMKM makanan di TikTok ke dalam kategori positif, negatif, dan netral menggunakan algoritma Multinomial Naive Bayes. Dataset terdiri dari 1.041 data komentar bersih yang dikumpulkan melalui web scraping. Tahapan text preprocessing meliputi case folding, normalisasi slang, stopword removal, serta pembersihan data kosong dan duplikat. Pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based sedangkan pembobotan fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Dataset dibagi menggunakan rasio 80% data latih (832 data) dan 20% data uji (209 data). Hasil evaluasi performa model menunjukkan bahwa algoritma Naive Bayes mencapai tingkat akurasi sebesar 75,60%. Model sangat responsif pada kelas netral dengan F1-score sebesar 0,85 dan presisi kelas positif sebesar 1,00. Penelitian ini memberikan kontribusi praktis bagi pelaku UMKM dalam mengevaluasi tanggapan konsumen serta menyusun strategi pemasaran digital yang lebih responsif.

ABSTRACT

This The rapid growth of TikTok social media has made it an effective digital promotion medium for Micro, Small, and Medium Enterprises (MSME). —The large number of comments left by users serves as a valuable source of information to measure customer perception and satisfaction. However, the characteristics of social media text, which are unstructured, informal, and contain slang and emojis, make manual analysis difficult. This study aims to classify customer review sentiments regarding MSME food products on TikTok into positive, negative, and neutral categories using the Multinomial Naive Bayes algorithm. The dataset consists of 1,041 clean comment entries collected via web scraping. Text preprocessing steps include case folding, slang normalization, stopword removal, and the removal of empty and duplicate data. Sentiment labeling was conducted using a lexicon-based approach, while feature weighting was performed using Term Frequency-Inverse Document Frequency (TF-IDF). The dataset was split into an 80% training set (832 data points) and a 20% testing set (209 data points). The model performance evaluation showed that the Naive Bayes algorithm achieved an overall accuracy rate of 75.60%. The model demonstrated high performance on the neutral class with an F1-score of 0.85 and a precision of 1.00 for the positive class. This research provides a practical contribution for MSME owners in evaluating customer feedback and developing more responsive digital marketing strategies.

PENDAHULUAN

Perubahan signifikan telah terjadi di berbagai aspek kehidupan, termasuk dunia bisnis dan pemasaran, akibat perkembangan teknologi informasi dan internet. Karena media sosial dianggap mampu menjangkau pelanggan secara cepat, luas, dan efektif, penggunaannya untuk tujuan

promosi pun semakin meningkat. TikTok merupakan salah satu platform media sosial yang berkembang pesat. Selain menjadi sarana hiburan, TikTok kini juga telah dikembangkan menjadi alat pemasaran digital bagi perusahaan, khususnya Usaha Mikro, Kecil, dan Menengah (UMKM) (Lifinda Ayuning Putri dkk., 2024)

Permasalahan yang dihadapi adalah banyaknya komentar pelanggan pada media sosial TikTok yang sulit dianalisis secara manual oleh pelaku UMKM. Karakteristik teks tidak terstruktur seperti kesalahan penulisan, penggunaan emotikon, bahasa tidak baku, dan singkatan juga ditemukan dalam komentar TikTok, sehingga menyulitkan analisis sentimen. Masih sedikit penelitian yang membahas persepsi pengguna TikTok terhadap produk UMKM, dan penelitian yang ada umumnya hanya menggunakan akurasi sebagai tolok ukur keberhasilan. Dengan memanfaatkan permasalahan tersebut. Proses analisis mencakup tahap pra-pemrosesan data meliputi stemming, case folding, tokenisasi, dan penghapusan stopword guna memastikan data tersusun dengan rapi. Selain itu, evaluasi model dilakukan menggunakan confusion matrix yang mencakup metrik akurasi, presisi, recall, dan skor F1, sehingga memungkinkan penilaian menyeluruh terhadap hasil klasifikasi. Berbagai algoritma pembelajaran mesin telah dikembangkan dan diterapkan dalam upaya untuk menafsirkan data sentimen secara otomatis dan efektif. Salah satu algoritma paling populer yang bekerja dengan baik untuk analisis sentimen dan tugas kategorisasi teks lainnya, adalah algoritma Naive Bayes (Made Dwi Utami Putra dkk., 2025; Puspitasari & Annisa, 2025). Berdasarkan hal tersebut, penelitian ini bertujuan untuk menggunakan metode Naive Bayes guna menganalisis sentimen konsumen terhadap produk UMKM di TikTok. Hasil penelitian ini diharapkan dapat membantu pemilik UMKM memahami sikap konsumen, yang pada gilirannya menjadi landasan bagi penilaian produk, peningkatan layanan, serta pengembangan strategi pemasaran digital yang lebih efektif.

Berdasarkan penelitian terkini, algoritma Naive Bayes sering digunakan dalam analisis sentimen karena tingkat akurasinya yang memadai dan kecepatan pemrosesan komputasi yang relatif tinggi. Algoritma Naive Bayes digunakan dalam penelitian oleh (Fadli & Hasan, 2024) untuk menganalisis sentimen terkait fitur-fitur TikTok Shop; hasilnya menunjukkan bahwa pendekatan tersebut mampu mengategorikan sentimen secara akurat berdasarkan komentar pengguna di TikTok Shop. Penelitian lain oleh (Tanggraeni & Sitokdana, 2022) menggunakan algoritma Naive Bayes untuk mengevaluasi sentimen terhadap aplikasi e-government di Google Play, dan temuannya membuktikan efektivitas metode tersebut dalam mengategorikan penilaian pengguna. Selain itu, studi oleh (Puspitasari & Annisa, 2025) menggunakan algoritma Naive Bayes untuk menganalisis sentimen pada takarir (caption) "Instagram for Business", yang membuktikan efektivitas metode ini dalam klasifikasi teks media sosial.

LANDASAN TEORI

Analisis Sentimen

Tujuan dari analisis sentimen sebuah bidang dalam text mining dan pemrosesan bahasa alami (NLP) adalah untuk mengenali, menganalisis, dan mengategorikan keyakinan, perasaan, atau sikap yang terkandung dalam data berbasis teks (Islam & Shahin, 2023). Berdasarkan isi teks yang diperiksa, metode ini digunakan untuk mengklasifikasikan opini sebagai netral, negatif, atau positif. Berkat kemajuan teknologi analisis informasi, analisis sentimen sering digunakan untuk mengekstrak informasi yang berguna bagi pengambilan keputusan dari data yang berasal dari media sosial, forum diskusi, situs ulasan, dan platform e-commerce (Hutabarat dkk., 2024)

Usaha, Mikro, Kecil dan Menengah (UMKM)

Usaha mikro, kecil, dan menengah memegang peranan penting dalam pertumbuhan ekonomi Indonesia. Sektor ini berkontribusi terhadap perekonomian melalui penciptaan lapangan kerja, penerimaan pajak bagi negara, serta pemerataan kesejahteraan yang lebih baik. Usaha Kecil dan Menengah (UKM) memiliki potensi pengembangan yang besar di era digital berkat kemampuan

adaptasi yang luar biasa terhadap perubahan kondisi ekonomi. Sebagaimana dinyatakan oleh Badan Pusat Statistik (BPS), UKM memainkan peran penting dalam perekonomian daerah dan memberikan kontribusi signifikan terhadap perekonomian nasional.

Algoritma Naive Bayes

Berdasarkan teorema Bayes, algoritma Naive Bayes merupakan metode klasifikasi probabilistik yang mengasumsikan independensi fitur, atau sifat naive (Ardiani dkk., 2020). Berkat kemudahan penggunaan, efektivitas, dan kinerjanya yang sering kali kompetitif bahkan jika dibandingkan dengan model-model yang lebih canggih algoritma ini banyak digunakan dalam berbagai aplikasi pembelajaran mesin, khususnya untuk tugas kategorisasi teks (Tanggraeni & Sitokdana, 2022).

Media Sosial Tiktok

Dalam TikTok adalah situs web media sosial yang memungkinkan pengguna membuat, membagikan, dan menonton video pendek. Platform ini telah berkembang pesat menjadi jaringan media sosial yang signifikan dengan basis pengguna yang besar, khususnya di kalangan anak muda. Saat ini, selain digunakan untuk hiburan, TikTok juga berfungsi sebagai sarana pemasaran digital dan bisnis.

Text Mining

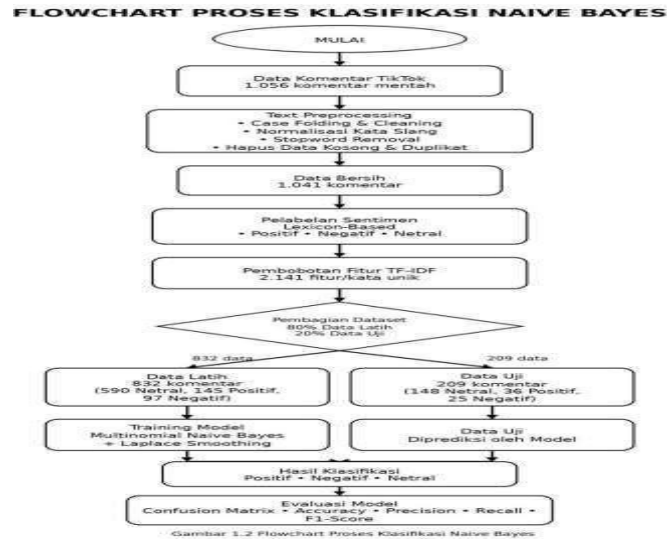
Perhitungan Natural Language Processing (NLP) dan penambangan data (data mining) melahirkan penambangan teks (text mining), sebuah subbidang yang berfokus pada ekstraksi pengetahuan dari data berbasis teks. Tujuannya adalah mengorganisasi informasi tekstual agar dapat dianalisis menggunakan berbagai metode komputasi. Menurut Feldman dan Sanger (2007), proses menganalisis dokumen teks secara otomatis untuk menemukan informasi yang sebelumnya tidak diketahui disebut sebagai penambangan teks. Volume data teks yang dihasilkan pengguna telah meningkat secara signifikan akibat perkembangan media sosial dan internet. Komentar, ulasan, opini, artikel, dan diskusi di berbagai media digital menjadi sumber data ini. Data teks tersebut umumnya memuat informasi tidak terstruktur yang perlu diproses sebelum dapat dijadikan dasar pengambilan keputusan.

Test Preprocessing

Flask Sebelum dianalisis menggunakan algoritma klasifikasi, input teks yang tidak terstruktur harus dibersihkan, dinormalisasi, dan ditransformasikan melalui tahap pra-pemrosesan teks dalam proses text mining. Mengingat adanya bahasa non-standar, tanda baca, simbol, angka, emotikon, dan URL dalam data media sosial, pra-pemrosesan menjadi langkah krusial untuk meningkatkan kualitas data dan menjamin analisis yang lebih akurat. Dalam text mining, pra-pemrosesan teks sangat penting untuk mengurangi noise pada data masukan serta mempersiapkan data untuk tahap ekstraksi fitur dan klasifikasi (Feldman dan Sanger, 2007).

METODE PENELITIAN

Penelitian ini dilakukan secara mandiri tanpa terikat pada instansi atau lembaga lainnya. Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap produk UMKM makanan di TikTok ke dalam kategori positif, negatif, dan netral menggunakan algoritma Multinomial Naive Bayes. Dataset terdiri dari 1.041 data komentar bersih yang dikumpulkan melalui web scraping. Tahapan text preprocessing meliputi case folding, normalisasi slang, stopword removal, serta pembersihan data kosong dan duplikat. Pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based, sedangkan pembobotan fitur menggunakan Term Frequency-Inverse Document Frequency (TF-IDF). Dataset dibagi menggunakan rasio 80% data latih (832 data) dan 20% data uji (209 data).



Gambar 1 Flowchart Proses Klasifikasi Naive Bayes

HASIL DAN PEMBAHASAN

Hasil

Hasil evaluasi performa model menunjukkan bahwa algoritma Naive Bayes mencapai tingkat akurasi sebesar 75,60%. Model sangat responsif pada kelas netral dengan F1-score sebesar 0,85 dan presisi kelas positif sebesar 1,00. Penelitian ini memberikan kontribusi praktis bagi pelaku UMKM dalam mengevaluasi tanggapan konsumen serta menyusun strategi pemasaran digital yang lebih responsif.

Pengujian dan Evaluasi Performa Model Naive Bayes

Model Multinomial Naive Bayes, yang menghitung kemungkinan kemunculan kata dalam setiap kategori sentimen, digunakan dalam penelitian ini. Pengujian dilakukan terhadap 20% data uji (209 baris data) untuk melihat sejauh mana model mampu menebak sentimen secara akurat berdasarkan fitur TF-IDF yang telah diekstraksi.

Precision, Recall, F1-Score, dan Accuracy adalah empat metrik utama yang digunakan untuk menilai model evaluasi kinerja. Tabel 4.6 di bawah ini menampilkan hasil pengujian klasifikasi Multinomial Naive Bayes:

Tabel 1. Hasil Evaluasi Performa Klasifikasi Naive Bayes

Kelas Sentimen	Precision	Recall	F1-Score	Support
Positif	1,00	0,28	0,43	36
Negatif	0,00	0,00	0,00	25
Netral	0,74	1,00	0,85	148
Macro Avg	0,58	0,43	0,43	209
Weighted Avg	0,70	0,76	0,68	209
Akurasi	-	-	75,60%	209

Berdasarkan hasil pengujian pada Tabel 4.6 di atas, analisis terhadap performa model Multinomial Naive Bayes dapat dijabarkan sebagai berikut:

1. Akurasi Sistem: Model Multinomial Naive Bayes berhasil meraih tingkat akurasi sebesar 75,60% pada data uji.
2. Kinerja Kelas Netral: Memiliki nilai *Recall* sebesar 1,00 (100%) dan *F1-Score* 0,85, menunjukkan bahwa model sangat kuat dan akurat dalam mengenali komentar berlabel Netral karena proporsi data Netral mendominasi dataset (sebesar 70,9%).
3. Kinerja Kelas Positif: Memiliki nilai *Precision* sebesar 1,00 (100%) yang berarti seluruh komentar yang diprediksi sebagai Positif oleh model adalah tepat, meskipun nilai *Recall*-nya berada pada 0,28 (sebagian komentar positif diprediksi ke kelas mayoritas/netral).
4. Kinerja Kelas Negatif: Memiliki nilai *Precision* dan *Recall* 0,00 dikarenakan jumlah data latih untuk kelas Negatif relatif terbatas (*imbalanced dataset*), sehingga model cenderung mengklasifikasikan dokumen ke kelas mayoritas (Netral).

Pembahasan Kualitatif dan Analisis Kata Kunci

Untuk memberikan interpretasi bisnis dan akademis yang lebih mendalam, dilakukan analisis terhadap kata-kata yang paling sering muncul (*frequent words*) pada masing-masing kelompok sentimen berdasarkan hasil ekstraksi kata kunci dari komentar bersih.

Pada kategori positif, kata kunci yang paling dominan muncul antara lain: semoga, nular, keren, masyaallah, sukses, pas, dan basreng. Kemunculan kata kunci ini mengindikasikan bahwa persepsi positif pelanggan sangat didorong oleh apresiasi terhadap kualitas produk (seperti kecocokan/kesesuaian rasa dan takaran produk *basreng*), respon emosional yang baik (pujian "*keren*" dan "*masyaallah*"), serta doa dan harapan positif pelanggan terhadap kesuksesan penjual di TikTok Shop ("*semoga nular*" dan "*sukses*").

Pada kategori negatif, kata kunci yang dominan adalah: ga, gak, mahal, tapi, beli, dan jualan. Hal ini menunjukkan bahwa ketidakpuasan pelanggan terutama dipicu oleh pertimbangan harga produk yang dianggap relatif mahal oleh sebagian pembeli, serta ungkapan keraguan atau kekecewaan yang disampaikan dengan kata konjungsi pertentangan ("*tapi*", "*ga*", "*gak*") terkait pengalaman saat membeli atau membandingkan produk dengan ekspektasi mereka.

Sedangkan pada kategori netral, kata kunci yang paling dominan muncul antara lain: kak, basreng, beli, jualan, apa, bumbu, bikin, dan semangat. Kemunculan kata kunci ini menunjukkan bahwa sebagian besar audiens atau penonton video TikTok menyampaikan komentar berupa pertanyaan umum mengenai informasi produk ("*apa nama bumbu/bahannya*", "*cara beli*", "*bikin dari apa*"), interaksi/sapaan kepada penjual ("*kak/ka/kk*"), serta ungkapan dukungan deskriptif atau pernyataan umum terkait aktivitas jualan tanpa memberikan penilaian emosional secara langsung (baik bernada sangat memuji maupun kecewa).

KESIMPULAN DAN SARAN

Kesimpulan

1. Penerapan Algoritma Naive Bayes: Algoritma *Multinomial Naive Bayes* berhasil diterapkan untuk mengklasifikasikan data komentar ulasan pelanggan produk UMKM di TikTok melalui tahapan *text preprocessing* (*case folding, normalisasi slang, stopword removal*), pelabelan berbasis *lexicon*, dan ekstraksi fitur TF-IDF.
2. Distribusi Sentimen Pelanggan: Dari total 1.041 data komentar bersih yang dianalisis, diperoleh distribusi sentimen mayoritas berupa kelas Netral sebanyak 738 komentar (70,9%), diikuti kelas Positif sebanyak 181 komentar (17,4%), dan kelas Negatif sebanyak 122 komentar (11,7%). Hal ini menunjukkan bahwa interaksi audiens TikTok terhadap produk UMKM makanan didominasi oleh pertanyaan seputar informasi produk, harga, dan cara pemesanan.
3. Akurasi dan Evaluasi Performa Model: Pengujian model menggunakan 209 data uji menghasilkan tingkat akurasi keseluruhan sebesar 75,60%. Model memiliki performa prediktif yang sangat tinggi pada kelas Positif dengan nilai *Precision* mencapai 1,00 (100%) dan performa seimbang pada kelas Netral (*F1-Score* 0,85).

Saran

1. Penanganan Data Tidak Seimbang (*Imbalanced Data*): Penelitian selanjutnya disarankan menerapkan teknik pemrosesan khusus untuk mengatasi ketidakseimbangan kelas, seperti *Resampling* (SMOTE/Random Oversampling) atau penyesuaian bobot kelas (*class weights*) agar presisi dan *recall* pada kelas minoritas (negatif) dapat ditingkatkan.
2. Pengembangan Teknik Preprocessing: Menambahkan kamus normalisasi kata *slang* khas media sosial yang lebih komprehensif serta penanganan emoji secara spesifik agar informasi emosional dari simbol visual dapat terkonversi menjadi nilai sentimen.

Perbandingan Metode: Mengintegrasikan atau membandingkan kinerja algoritma Naive Bayes dengan algoritma *Machine Learning* atau *Deep Learning* lainnya seperti *Support Vector Machine* (SVM), *Random Forest*, atau *BERT* untuk menguji peningkatan performa akurasi klasifikasi teks.

DAFTAR PUSTAKA

- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi Sentiment Analysis Tanggapan Masyarakat Terhadap Pembangunan di Kota Pontianak. *Jurnal Sistem dan Teknologi Informasi (Justin)*, 8(2), 183. <https://doi.org/10.26418/justin.v8i2.36776>
- Delicia, C., & Paramita, S. (2022). Konten Tiktok untuk Meningkatkan UMKM di Masa Pandemi Covid-19 (Studi Kasus Rude Basic dan Agate Deluxe). *Koneksi*, 6(1), 218. <https://doi.org/10.24912/kn.v6i1.15768>
- Fadli, K., & Hasan, F. N. (2024). Analisis Sentimen Tanggapan Masyarakat Terhadap Penutupan TikTok Shop Menggunakan Metode Naïve Bayes. *Journal of Information System Research (JOSH)*, 6(1), 396–405. <https://doi.org/10.47065/josh.v6i1.6060>
- 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>
- Hutabarat, A. W. V., Adnyani, N. L. S. S., & Suryadi, K. (2024). Analisis Sentimen Data Ulasan Pengguna MyPertamina di Twitter dengan Metode Text Mining. *Jurnal Rekayasa Sistem Industri*, 13(1), 145–154. <https://doi.org/10.26593/jrsi.v13i1.6958.145-154>
- Islam, M. J., & Shahin, M. B. N. (2023). Sentiment analysis of social media text using natural language processing. *World Journal of Advanced Research and Reviews*, 18(3), 1707–1714. <https://doi.org/10.30574/wjarr.2023.18.3.1161>
- Kurniawan, A. S., Fitriana, R., Vrisaliani, M., Saffa Adesti, N., & Alya Rahmadhani, S. (2025). *Pemanfaatan Media Sosial Tiktok untuk Pemasaran Bisnis Digital sebagai Media Promosi* (Vol. 11).
- Lestari, N. I., & Ramadani, M. (2022). *STRATEGI PEMASARAN PRODUK UMKM MELALUI PENERAPAN DIGITAL MARKETING (STUDI KASUS UMKM DI SERANG, BANTEN)*.
- Lifinda Ayuning Putri, I., Fadila Astri, M., Aulia Rachman, H., Dimar Alam, M., & Ekonomi dan Bisnis, F. (2024). *Sosialisasi Strategi Marketing Bisnis Melalui Media Sosial Tiktok*. 6, 286–293. <https://doi.org/10.55338/jpkmn.v6i1.1.4804>
- Made Dwi Utami Putra, D., Dharma Suryawan, W., Andayani, S., & Yunia Rusadi, E. (2025). *Analisis Sentimen terhadap Fenomena Artis Berjualan di Tiktok Shop pada Media Sosial X Menggunakan Model Naïve Bayes*.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. <https://doi.org/10.1017/CBO978051180907>
- Mardiana, H., Walid, M., & Kisnu Darmawan, A. (2023). *Lexicon-Based and Naïve Bayes Sentiment ... LEXICON-BASED AND NAIVE*