
Analisis Sentimen Konsumen Terhadap Usaha Food And Beverage Di Tembilahan Pada Media Sosial Tiktok Menggunakan Naïve Bayes Classifier Sebagai Dasar Evaluasi Strategi Pemasaran Digital

M. Saleh¹⁾, Dwi Yuli Prasetyo²⁾

^{1,2)}Sistem Informasi/Fakultas Teknik dan Ilmu Komputer, Universitas Islam Indragiri

Email : msaleh0205@gmail.com
dwiyuliprasetyo2@mail.com

Abstrak

Media sosial TikTok menjadi salah satu platform yang banyak digunakan konsumen untuk memberikan opini mengenai produk dan layanan, termasuk pada sektor Food and Beverage (F&B). Informasi yang terdapat pada komentar pengguna dapat dimanfaatkan untuk mengetahui persepsi konsumen dan mengevaluasi strategi pemasaran digital. Penelitian ini bertujuan menganalisis sentimen konsumen terhadap usaha F&B di Tembilahan pada media sosial TikTok menggunakan algoritma Naïve Bayes Classifier. Penelitian menggunakan pendekatan kuantitatif dengan data sebanyak 500 komentar TikTok yang diperoleh melalui teknik scraping. Data kemudian melalui tahap preprocessing, pelabelan sentimen, ekstraksi fitur menggunakan TF-IDF, dan klasifikasi menggunakan Naïve Bayes. Evaluasi model dilakukan menggunakan confusion matrix, accuracy, precision, recall, F1-score, serta 5-Fold Cross Validation. Hasil penelitian menunjukkan bahwa sentimen netral mendominasi dengan 248 komentar (49,6%), diikuti sentimen positif sebanyak 156 komentar (31,2%) dan sentimen negatif sebanyak 96 komentar (19,2%). Model Naïve Bayes menghasilkan akurasi sebesar 93,00% dengan rata-rata akurasi cross validation sebesar 92,00%. Hasil ini menunjukkan bahwa Naïve Bayes mampu mengklasifikasikan sentimen konsumen dengan baik dan dapat digunakan sebagai dasar evaluasi strategi pemasaran digital usaha F&B di Tembilahan

Kata kunci: Analisis Sentimen, TikTok, Food and Beverage, Naïve Bayes Classifier, Pemasaran Digital.

Abstract

TikTok has become one of the most widely used social media platforms where consumers share opinions and experiences regarding products and services, including those in the Food and Beverage (F&B) sector. User comments on TikTok can provide valuable insights into consumer perceptions and serve as a basis for evaluating digital marketing strategies. This study aims to analyze consumer sentiment toward F&B businesses in Tembilahan using the Naïve Bayes Classifier algorithm. This research employed a quantitative approach using 500 TikTok comments collected through a scraping process. The data underwent preprocessing, sentiment labeling, feature extraction using TF-IDF, and classification using Naïve Bayes. Model performance was evaluated using a confusion matrix, accuracy, precision, recall, F1-score, and 5-Fold Cross Validation. The results showed that neutral sentiment dominated with 248 comments (49.6%), followed by positive sentiment with 156 comments (31.2%) and negative sentiment with 96 comments (19.2%). The Naïve Bayes model achieved an accuracy of 93.00% and an average cross-validation accuracy of 92.00%. These findings indicate that the Naïve Bayes algorithm is effective in classifying consumer sentiment and can be utilized as a tool for evaluating digital marketing strategies of F&B businesses in Tembilahan.

Keywords: Sentiment Analysis, TikTok, Food and Beverage, Naïve Bayes Classifier, Digital Marketing.

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mengubah paradigma pemasaran dari pendekatan konvensional menjadi pemasaran digital yang lebih interaktif, terukur, dan berorientasi pada konsumen. Digitalisasi membuka peluang bagi pelaku usaha untuk membangun komunikasi dua arah dengan pelanggan melalui berbagai platform media sosial. Media sosial kini tidak hanya berfungsi sebagai sarana berbagi informasi, tetapi juga telah berkembang menjadi media strategis dalam membangun citra merek, meningkatkan keterlibatan konsumen (customer engagement), memperluas pangsa pasar, serta memengaruhi keputusan pembelian (Social & Meltwater, 2024). Perubahan ini mendorong pelaku usaha untuk

memanfaatkan data dari aktivitas pengguna media sosial sebagai sumber informasi bernilai dalam mendukung pengambilan keputusan bisnis berbasis data (data-driven decision making).

Salah satu platform media sosial dengan pertumbuhan paling pesat adalah TikTok. Sejak diluncurkan secara global, TikTok berkembang dari platform berbasis video pendek yang awalnya berfungsi sebagai hiburan menjadi media promosi, pemasaran, edukasi, dan komunikasi antara produsen dan konsumen. Berdasarkan laporan (DataReportal, 2025), Indonesia merupakan salah satu negara dengan jumlah pengguna TikTok terbesar di dunia, dengan lebih dari 126 juta pengguna berusia di atas 18 tahun. Tingginya penetrasi pengguna ini menjadikan TikTok sebagai salah satu media pemasaran digital paling potensial, khususnya bagi pelaku usaha Food and Beverage (F&B) yang mengandalkan penyajian visual produk, promosi kreatif, dan interaksi langsung dengan konsumen. Algoritma TikTok yang mampu merekomendasikan konten sesuai preferensi pengguna juga memberi peluang lebih besar bagi pelaku usaha lokal untuk menjangkau pasar yang lebih luas tanpa memerlukan biaya promosi yang tinggi.

Keberhasilan suatu strategi pemasaran digital tidak hanya diukur dari jumlah tayangan (views), pengikut (followers), atau tingkat interaksi (engagement rate), tetapi juga dipengaruhi oleh persepsi konsumen terhadap produk atau layanan yang ditawarkan. Persepsi tersebut umumnya tercermin melalui komentar, ulasan, dan tanggapan pengguna pada konten yang dipublikasikan, yang mengandung informasi tentang tingkat kepuasan, pengalaman penggunaan produk, kualitas pelayanan, harga, hingga efektivitas promosi. Menurut (Liu, 2022), opini pengguna yang dipublikasikan melalui media sosial merupakan bentuk electronic word of mouth (e-WOM) yang berpengaruh besar terhadap perilaku konsumen karena mampu membentuk persepsi masyarakat terhadap suatu produk atau merek. Oleh karena itu, analisis terhadap komentar pengguna menjadi penting untuk memahami kebutuhan dan harapan konsumen secara lebih objektif.

Besarnya volume komentar yang dihasilkan pengguna media sosial membuat analisis manual menjadi kurang efektif dan memakan waktu relatif lama. Kondisi ini mendorong berkembangnya text mining sebagai pendekatan untuk mengolah data teks secara otomatis, yaitu proses ekstraksi informasi dari kumpulan dokumen tidak terstruktur untuk menghasilkan pengetahuan yang dapat dimanfaatkan dalam pengambilan keputusan (Bird et al., 2021). Salah satu penerapan text mining yang banyak digunakan dalam pemasaran digital adalah analisis sentimen (sentiment analysis), yang bertujuan mengidentifikasi dan mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral berdasarkan teks yang dihasilkan (Liu, 2022). Teknik ini telah diterapkan secara luas, mulai dari evaluasi kualitas layanan, analisis kepuasan pelanggan, analisis ulasan produk, hingga prediksi perilaku konsumen di media sosial.

Dalam proses analisis sentimen, tahap preprocessing memegang peran penting untuk meningkatkan kualitas data sebelum dianalisis. Tahap ini meliputi case folding, cleaning, tokenizing, stopword removal, dan stemming, sehingga data teks menjadi lebih terstruktur dan siap diproses oleh algoritma klasifikasi (Tala, 2021). Setelah preprocessing, teks umumnya direpresentasikan menggunakan pembobotan kata seperti Term Frequency–Inverse Document Frequency (TF-IDF), yang memberikan bobot lebih tinggi pada kata-kata dengan tingkat diskriminasi tinggi dalam dokumen. Representasi data yang baik pada tahap ini akan meningkatkan performa algoritma klasifikasi dalam mengidentifikasi kecenderungan sentimen pengguna.

Berbagai algoritma machine learning telah dikembangkan untuk klasifikasi sentimen, di antaranya Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbor, Logistic Regression, hingga Naïve Bayes Classifier (NBC). Di antara algoritma tersebut, Naïve Bayes tetap menjadi salah satu metode yang banyak digunakan karena proses komputasinya sederhana, efisien, dan mampu menghasilkan akurasi yang baik pada data teks (Manning et al., 2022). Algoritma ini bekerja berdasarkan Teorema Bayes dengan asumsi independensi antarfitur, sehingga proses pembelajarannya tetap cepat meskipun menggunakan data dalam jumlah besar

(Mitchell, 2021). Kompleksitas komputasinya yang relatif rendah membuat Naïve Bayes sesuai diterapkan pada penelitian yang memanfaatkan data komentar media sosial berskala besar.

Beberapa penelitian terdahulu menunjukkan bahwa analisis sentimen menggunakan Naïve Bayes tetap memberikan performa yang kompetitif. (Tsiourlini et al., 2024) menganalisis komentar pengguna YouTube terhadap produk pangan dan menemukan bahwa analisis sentimen mampu menggambarkan persepsi konsumen terhadap kualitas produk serta preferensi terhadap merek tertentu, sekaligus menunjukkan bahwa Naïve Bayes tetap efektif diterapkan pada data teks yang telah melalui preprocessing. (Visalli, 2026) menambahkan bahwa analisis sentimen telah berkembang menjadi pendekatan penting dalam consumer science karena mampu mengekstraksi makna afektif dari opini konsumen yang dipublikasikan melalui media digital.

Penelitian mengenai analisis sentimen juga berkembang pesat di Indonesia. (Layuk, 2025) menunjukkan bahwa penerapan feature selection mampu meningkatkan performa Naïve Bayes dalam klasifikasi sentimen konsumen, menghasilkan akurasi yang lebih tinggi dibandingkan tanpa seleksi fitur. (Ghiffary & Syafrullah, 2024) membuktikan bahwa Naïve Bayes efektif mengklasifikasikan sentimen pengguna TikTok Shop berdasarkan komentar pengguna. (Putra, 2025) menunjukkan bahwa Naïve Bayes dapat digunakan untuk menganalisis persepsi masyarakat terhadap fenomena bisnis digital di media sosial dengan akurasi yang baik, sementara (Kurniawan, 2025), melalui kajian sistematis mengenai analisis sentimen, menegaskan bahwa pemanfaatan data media sosial semakin penting sebagai sumber informasi dalam mendukung strategi pemasaran berbasis data.

Meskipun demikian, kajian literatur menunjukkan bahwa sebagian besar penelitian terdahulu masih berfokus pada perbandingan performa algoritma klasifikasi, peningkatan akurasi model, atau analisis sentimen pada platform e-commerce, layanan publik, dan isu nasional. Penelitian yang secara khusus membahas sentimen konsumen terhadap usaha F&B lokal melalui komentar TikTok masih relatif terbatas. Sebagian besar penelitian juga berhenti pada evaluasi performa model klasifikasi, tanpa menghubungkan hasil analisis sentimen dengan penyusunan rekomendasi strategi pemasaran digital yang dapat langsung diterapkan oleh pelaku usaha. Kondisi ini menunjukkan adanya kesenjangan penelitian (research gap) yang perlu dikaji lebih lanjut.

Fenomena tersebut juga ditemukan pada berbagai usaha F&B di Kota Tembilahan, Kabupaten Indragiri Hilir. Dalam beberapa tahun terakhir, pelaku usaha semakin aktif memanfaatkan TikTok sebagai media promosi melalui konten video pendek, ulasan produk, testimoni pelanggan, informasi promosi, maupun peluncuran menu baru. Namun, evaluasi terhadap tanggapan konsumen masih dilakukan secara manual, sehingga informasi yang diperoleh cenderung subjektif dan belum menggambarkan persepsi konsumen secara menyeluruh — padahal komentar pengguna merupakan sumber informasi berharga untuk mengetahui kepuasan pelanggan, kualitas pelayanan, persepsi harga, serta efektivitas strategi pemasaran yang diterapkan.

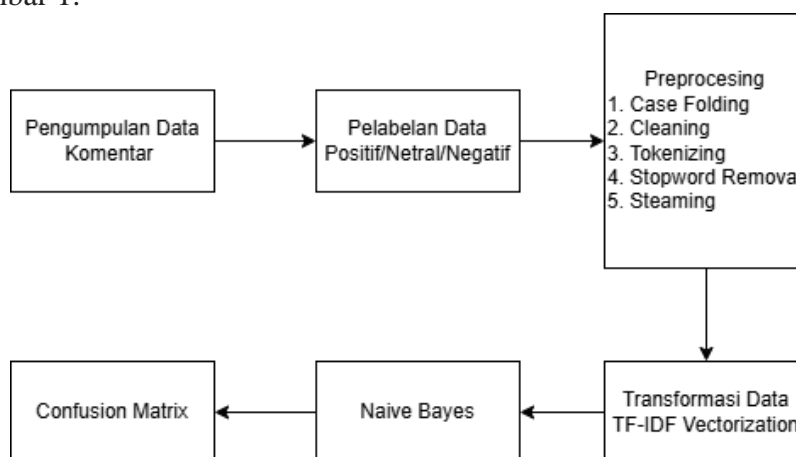
Berdasarkan uraian tersebut, penelitian ini menawarkan kebaruan (novelty) berupa penerapan analisis sentimen menggunakan Naïve Bayes Classifier pada komentar pengguna TikTok untuk menganalisis persepsi konsumen terhadap usaha F&B lokal di Kota Tembilahan. Berbeda dari penelitian sebelumnya yang berorientasi pada peningkatan performa algoritma, penelitian ini menghubungkan hasil klasifikasi sentimen dengan evaluasi strategi pemasaran digital, sehingga memberikan manfaat yang lebih aplikatif bagi pelaku usaha. Dengan demikian, penelitian ini diharapkan berkontribusi tidak hanya pada pengembangan metode analisis sentimen berbasis machine learning, tetapi juga pada penyusunan strategi pemasaran digital berbasis data bagi usaha F&B lokal.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan menganalisis sentimen konsumen terhadap usaha Food and Beverage di Kota Tembilahan berdasarkan komentar pada media sosial TikTok menggunakan metode Naïve Bayes Classifier. Secara khusus, penelitian ini bertujuan mengidentifikasi distribusi sentimen konsumen, mengevaluasi kinerja Naïve Bayes

dalam klasifikasi sentimen, serta menghasilkan rekomendasi strategi pemasaran digital berdasarkan hasil analisis sentimen yang diperoleh. Hasil penelitian diharapkan menjadi referensi bagi pelaku usaha dalam memanfaatkan data media sosial sebagai dasar pengambilan keputusan bisnis, sekaligus memperkaya kajian penerapan text mining dan machine learning dalam bidang pemasaran digital.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik text mining dan algoritma Naïve Bayes Classifier (NBC) untuk menganalisis sentimen konsumen terhadap usaha Food and Beverage (F&B) di Kota Tembilahan berdasarkan komentar pengguna pada media sosial TikTok. Pendekatan kuantitatif dipilih karena memungkinkan pengukuran dan analisis data secara objektif menggunakan metode statistik dan komputasi (Creswell & Creswell, 2023; Sekaran & Bougie, 2020). Tahapan penelitian meliputi pengumpulan data, pelabelan sentimen, preprocessing, ekstraksi fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), pembentukan model klasifikasi Naïve Bayes, serta evaluasi performa model menggunakan hold-out validation dan 5-Fold Cross Validation. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

1. Pengumpulan Data

Data yang digunakan merupakan data sekunder berupa komentar pengguna TikTok yang diperoleh dari akun atau konten usaha F&B di Kota Tembilahan melalui teknik web scraping. Teknik ini dipilih karena mampu mengumpulkan data media sosial secara otomatis dalam jumlah besar, sehingga lebih efisien dibandingkan pengumpulan data manual. Komentar yang dikumpulkan mencakup konten terkait promosi produk, kualitas makanan dan minuman, pelayanan, harga, maupun pengalaman pelanggan. Seluruh data disimpan dalam format Comma Separated Values (.csv) untuk memudahkan pengolahan menggunakan bahasa pemrograman Python.

Teknik pengambilan sampel menggunakan purposive sampling, yaitu memilih komentar yang memenuhi kriteria tertentu: menggunakan bahasa Indonesia, bukan komentar spam, dan memiliki keterkaitan dengan persepsi konsumen terhadap usaha F&B. Teknik ini dipilih agar data yang digunakan benar-benar relevan dengan tujuan penelitian (Taherdoost, 2020).

2. Pelabelan Data

Tahap selanjutnya adalah pelabelan (labeling) data untuk menentukan kategori sentimen setiap komentar ke dalam tiga kelas: positif, negatif, dan netral. Sentimen positif menunjukkan kepuasan atau apresiasi konsumen terhadap produk maupun pelayanan; sentimen negatif

menunjukkan kritik atau ketidakpuasan; sedangkan sentimen netral diberikan pada komentar yang tidak menunjukkan kecenderungan opini tertentu (Liu, 2022).

Pelabelan dilakukan secara manual oleh peneliti berdasarkan konteks kalimat pada setiap komentar. Untuk meningkatkan konsistensi hasil pelabelan, sebagian data diverifikasi oleh validator yang memahami konsep analisis sentimen, sehingga mengurangi kemungkinan kesalahan klasifikasi pada data latih.

3. Preprocessing Data

Sebelum dilakukan proses klasifikasi, seluruh komentar melalui tahapan preprocessing untuk membersihkan data dari unsur-unsur yang tidak diperlukan. Tahapan ini bertujuan meningkatkan kualitas data sehingga menghasilkan representasi teks yang lebih baik (Bird et al., 2021). Tahapan preprocessing meliputi:

- a) Case Folding, yaitu mengubah seluruh huruf menjadi huruf kecil (lowercase).
- b) Cleaning, yaitu menghapus URL, emoji, simbol, angka, tanda baca, dan karakter khusus lainnya.
- c) Tokenizing, yaitu memecah kalimat menjadi kumpulan kata (tokens).
- d) Stopword Removal, yaitu menghapus kata-kata umum yang tidak memiliki makna penting dalam proses klasifikasi.
- e) Stemming, yaitu mengubah kata berimbuhan menjadi bentuk dasar menggunakan pustaka Sastrawi (Tala, 2021).

Melalui tahapan ini, data komentar menjadi lebih bersih dan siap digunakan dalam proses ekstraksi fitur.

4. Ekstraksi Fitur Menggunakan TF-IDF

Data hasil preprocessing kemudian diubah ke dalam bentuk numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF merupakan teknik pembobotan kata yang memberikan bobot berdasarkan frekuensi kemunculan kata dalam suatu dokumen dan tingkat kemunculannya pada seluruh dokumen. Semakin sering suatu kata muncul pada satu dokumen namun jarang muncul pada dokumen lain, maka bobot kata tersebut akan semakin tinggi. Representasi TF-IDF menghasilkan matriks numerik yang digunakan sebagai masukan (input) bagi algoritma Naïve Bayes dalam proses klasifikasi sentimen. Menggunakan persamaan (1) dan (2) adalah sebagai berikut:

$$TFIDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

5. Pembagian Dataset

Dataset yang telah ditransformasikan menggunakan TF-IDF dibagi menjadi data latih (training data) dan data uji (testing data) menggunakan metode hold-out validation dengan perbandingan 80% : 20%. Sebanyak 80% data digunakan untuk membangun model klasifikasi, sedangkan 20% sisanya digunakan untuk menguji kemampuan model dalam mengklasifikasikan data yang belum pernah dipelajari sebelumnya. Menurut (Géron, 2022), rasio 80:20 merupakan salah satu skema pembagian data yang umum digunakan dalam penelitian machine learning karena mampu memberikan keseimbangan antara proses pelatihan dan evaluasi model.

6. Klasifikasi Naive Bayes

Metode klasifikasi yang digunakan dalam penelitian ini adalah Multinomial Naïve Bayes Classifier. Algoritma Naïve Bayes merupakan metode klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen terhadap fitur lainnya (Mitchell, 2021). Naïve Bayes bekerja berdasarkan Teorema Bayes yang dinyatakan berdasarkan persamaan (3) sebagai berikut:

$$P(C|X) = \frac{P(X|C) \times P(C)}{P(X)} \quad (3)$$

Keterangan:

$P(C|X)$ = probabilitas kelas terhadap data

$P(X|C)$ = probabilitas data pada kelas tertentu

$P(C)$ = probabilitas awal kelas

$P(X)$ = probabilitas data

Naïve Bayes Classifier dipilih karena memiliki keunggulan dalam menangani data teks berdimensi tinggi, proses komputasi yang relatif cepat, serta mampu menghasilkan performa klasifikasi yang baik pada berbagai penelitian analisis sentimen (Manning et al., 2022). Selain itu, algoritma ini memiliki kompleksitas komputasi yang rendah dan efektif digunakan pada proses klasifikasi dokumen teks berbasis representasi fitur seperti TF-IDF (Mitchell, 2021). Penerapan metode Naïve Bayes yang dikombinasikan dengan feature selection mampu meningkatkan akurasi klasifikasi sentimen konsumen. Model Naïve Bayes dibangun menggunakan data latih (training data) untuk mempelajari pola hubungan antara fitur teks dan kategori sentimen. Selanjutnya, model yang telah terbentuk digunakan untuk memprediksi kategori sentimen pada data uji (testing data) ke dalam tiga kelas, yaitu positif, negatif, dan netral. Kinerja model kemudian dievaluasi menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score untuk mengukur kemampuan model dalam melakukan klasifikasi secara akurat.

7. Evaluasi Model

Evaluasi model dilakukan menggunakan Confusion Matrix untuk mengetahui tingkat performa klasifikasi yang dihasilkan. Metrik evaluasi yang digunakan dinyatakan berdasarkan persamaan (4), (5), (6), dan (7) sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4),$$

$$Precision = \frac{TP}{TP+FP} \quad (5),$$

$$Recall = \frac{TP}{TP+FN} \quad (6),$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7),$$

Keterangan:

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Menurut (Géron, 2022), accuracy, precision, recall, dan F1-score merupakan metrik evaluasi yang paling umum digunakan dalam penelitian klasifikasi karena mampu memberikan gambaran yang komprehensif mengenai kinerja model dalam melakukan prediksi.

Nilai accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data yang diuji. Sementara itu, precision mengukur tingkat ketepatan model dalam mengidentifikasi data pada suatu kelas, recall menunjukkan kemampuan model dalam menemukan seluruh data yang benar pada kelas tersebut, sedangkan F1-score merupakan rata-rata harmonis antara precision dan recall sehingga memberikan ukuran evaluasi yang lebih seimbang, terutama ketika distribusi data antar kelas tidak seimbang (Géron, 2022).

HASIL DAN PEMBAHASAN

1. Hasil

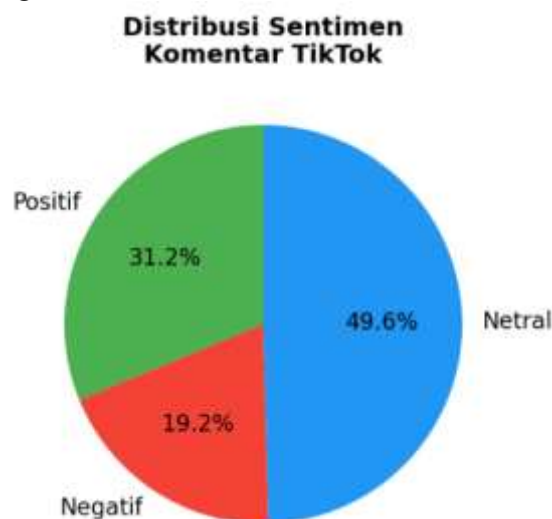
A. Pengumpulan Data

Data penelitian diperoleh dari komentar pengguna pada media sosial TikTok yang berkaitan dengan usaha Food and Beverage (F&B) di Tembilahan. Pengumpulan data dilakukan menggunakan teknik scraping komentar pada akun TikTok yang aktif mempublikasikan konten

promosi produk dan memiliki tingkat interaksi yang tinggi. Teknik ini dipilih karena mampu memperoleh data secara cepat dan dalam jumlah besar sehingga dapat merepresentasikan persepsi konsumen terhadap produk dan layanan yang ditawarkan oleh pelaku usaha. Komentar yang dikumpulkan mencakup berbagai bentuk tanggapan pengguna, seperti ulasan produk, pengalaman pembelian, kritik, saran, serta penilaian terhadap kualitas layanan. Pada tahap awal, seluruh komentar hasil scraping dikumpulkan dalam bentuk dataset mentah (raw data). Selanjutnya dilakukan proses seleksi data untuk menghapus komentar yang tidak relevan, data duplikat, spam, tautan (URL), simbol yang berlebihan, serta komentar yang tidak mengandung informasi sentimen. Proses ini bertujuan untuk meningkatkan kualitas dataset sehingga data yang digunakan benar-benar mencerminkan opini konsumen terhadap usaha F&B di Tembilahan. Setelah melalui tahap seleksi dan pembersihan data, diperoleh sebanyak 500 komentar yang digunakan sebagai dataset penelitian.

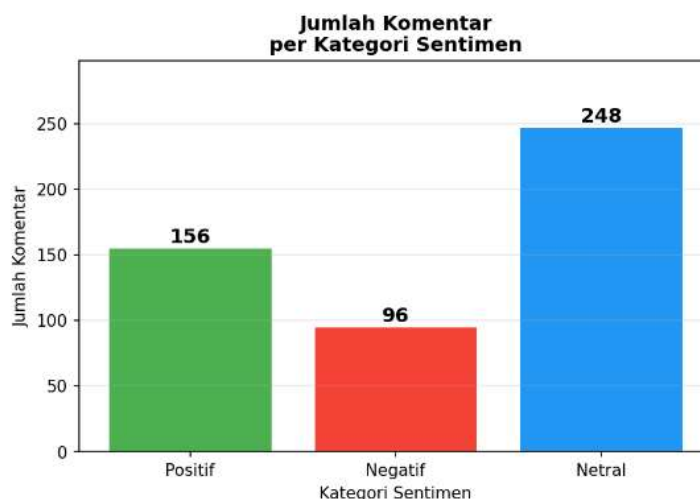
Data yang telah dikumpulkan kemudian diberi label sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Hasil pelabelan menunjukkan bahwa terdapat 156 komentar positif (31,2%), 96 komentar negatif (19,2%), dan 248 komentar netral (49,6%). Distribusi sentimen tersebut menunjukkan bahwa mayoritas pengguna TikTok memberikan komentar yang bersifat netral, sedangkan sentimen positif memiliki jumlah yang lebih tinggi dibandingkan sentimen negatif. Temuan ini mengindikasikan bahwa secara umum persepsi konsumen terhadap usaha F&B di Tembilahan cenderung positif, meskipun masih terdapat sejumlah komentar negatif yang dapat dijadikan bahan evaluasi bagi pelaku usaha.

Distribusi sentimen hasil pelabelan dapat dilihat pada Gambar 2. Berdasarkan visualisasi tersebut, sentimen netral mendominasi percakapan pengguna dengan persentase hampir setengah dari total komentar yang dianalisis. Dominasi sentimen netral menunjukkan bahwa sebagian besar pengguna menggunakan kolom komentar untuk memberikan tanggapan umum, bertanya mengenai produk, atau berinteraksi tanpa menunjukkan kecenderungan opini yang kuat terhadap produk maupun layanan yang ditawarkan.



Gambar 2. Distribusi Sentimen Komentar

Selain dalam bentuk persentase, distribusi sentimen juga ditampilkan berdasarkan jumlah komentar pada masing-masing kategori sebagaimana ditunjukkan pada Gambar 3. Visualisasi tersebut memperlihatkan bahwa kategori netral memiliki jumlah komentar tertinggi yaitu 248 komentar, diikuti sentimen positif sebanyak 156 komentar dan sentimen negatif sebanyak 96 komentar. Perbedaan jumlah komentar pada masing-masing kategori memberikan gambaran awal mengenai kecenderungan persepsi konsumen terhadap usaha F&B yang menjadi objek penelitian.



Gambar 3. Jumlah Komentar Berdasarkan Kategori

Hasil distribusi sentimen ini menjadi dasar untuk tahapan analisis berikutnya, yaitu preprocessing data, pembobotan fitur menggunakan TF-IDF, serta proses klasifikasi sentimen menggunakan algoritma Naïve Bayes Classifier. Melalui tahapan tersebut, penelitian berupaya mengidentifikasi pola sentimen konsumen secara lebih mendalam sehingga dapat digunakan sebagai dasar evaluasi strategi pemasaran digital usaha F&B di Tembilahan.

Hasil Preprocessing Data

Tahap preprocessing merupakan proses penting dalam analisis sentimen karena bertujuan untuk membersihkan dan menyiapkan data teks sebelum dilakukan proses klasifikasi. Komentar yang diperoleh dari TikTok umumnya masih mengandung berbagai karakter yang tidak terstruktur, seperti huruf kapital, tanda baca, emoji, angka, tautan, singkatan, dan kata-kata yang tidak memiliki makna penting dalam proses analisis. Oleh karena itu, preprocessing dilakukan untuk meningkatkan kualitas data sehingga dapat menghasilkan model klasifikasi yang lebih akurat.

Pada penelitian ini, preprocessing dilakukan melalui beberapa tahapan, yaitu case folding, cleaning, tokenizing, stopword removal, dan stemming. Tahap case folding digunakan untuk mengubah seluruh huruf menjadi huruf kecil (lowercase) sehingga tidak terjadi perbedaan antara kata yang memiliki makna sama tetapi ditulis dengan bentuk huruf yang berbeda. Selanjutnya, tahap cleaning dilakukan untuk menghapus karakter khusus, angka, emoji, tanda baca, tautan (URL), dan simbol lain yang tidak diperlukan dalam proses analisis sentimen.

Setelah proses cleaning, dilakukan tokenizing untuk memecah kalimat menjadi kumpulan kata (token) yang dapat diproses oleh sistem. Tahap berikutnya adalah stopword removal, yaitu menghapus kata-kata umum yang sering muncul namun tidak memberikan kontribusi signifikan terhadap klasifikasi sentimen, seperti "dan", "yang", "di", "ke", serta kata penghubung lainnya. Selanjutnya dilakukan proses stemming menggunakan kamus bahasa Indonesia untuk mengubah kata berimbuhan menjadi kata dasar sehingga variasi kata yang memiliki makna sama dapat direpresentasikan dalam bentuk yang seragam.

Contoh hasil preprocessing dapat dilihat pada Tabel 1. Tabel tersebut menunjukkan perubahan data komentar dari bentuk teks asli hingga menjadi data yang siap digunakan dalam proses pembobotan TF-IDF dan klasifikasi menggunakan algoritma Naïve Bayes Classifier.

Tabel 1. Contoh Hasil Preprocessing Data

Tahapan	Hasil
Data Asli	"Makanan nya enak bangettt □ wajib coba nih!"
Case Folding	makanan nya enak bangettt wajib coba nih
Cleaning	makanan nya enak bangettt wajib coba nih

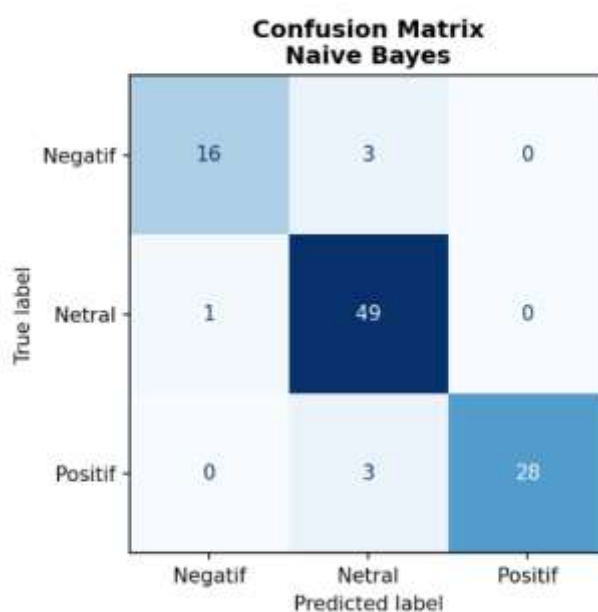
Tahapan	Hasil
Tokenizing	[makanan, nya, enak, bangettt, wajib, coba, nih]
Stopword Removal	[makanan, enak, bangettt, wajib, coba]
Stemming	[makan, enak, banget, wajib, coba]

Melalui proses preprocessing, seluruh komentar yang telah lolos tahap seleksi berhasil diubah menjadi data teks yang lebih terstruktur dan siap digunakan pada tahap pembobotan TF-IDF. Tahapan ini berperan penting dalam meningkatkan kualitas fitur yang digunakan oleh algoritma Naïve Bayes sehingga mampu menghasilkan performa klasifikasi yang lebih baik. Data hasil preprocessing kemudian digunakan sebagai masukan dalam proses ekstraksi fitur dan klasifikasi sentimen untuk mengidentifikasi kecenderungan opini konsumen terhadap usaha food and beverage di Tembilahan

Hasil Klasifikasi Naïve Bayes Classifier

Setelah proses preprocessing dan pembobotan fitur menggunakan TF-IDF selesai dilakukan, tahap selanjutnya adalah klasifikasi sentimen menggunakan algoritma Multinomial Naïve Bayes. Dataset yang digunakan dalam penelitian ini berjumlah 500 komentar TikTok yang telah diberi label sentimen, terdiri atas 156 komentar positif, 96 komentar negatif, dan 248 komentar netral. Dataset kemudian dibagi menjadi data latih (training data) sebesar 80% dan data uji (testing data) sebesar 20%. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi performa model dalam mengklasifikasikan sentimen komentar.

Evaluasi model dilakukan menggunakan confusion matrix untuk mengetahui tingkat ketepatan prediksi pada masing-masing kategori sentimen. Hasil confusion matrix ditunjukkan pada Gambar 4.



Gambar 4. Confusion Matrix Naïve Bayes

Berdasarkan Gambar 4, model berhasil mengklasifikasikan 93 dari 100 data uji dengan benar. Pada kategori sentimen negatif, sebanyak 16 komentar berhasil diprediksi dengan tepat sebagai sentimen negatif, sedangkan 3 komentar lainnya salah diklasifikasikan sebagai sentimen netral. Pada kategori sentimen netral, model mampu mengidentifikasi 49 komentar secara benar dan hanya 1 komentar yang salah diprediksi sebagai sentimen negatif. Sementara itu, pada

kategori sentimen positif, sebanyak 28 komentar berhasil diklasifikasikan dengan benar dan 3 komentar lainnya salah diprediksi sebagai sentimen netral.

Untuk mengukur performa model secara lebih komprehensif, digunakan metrik evaluasi berupa accuracy, precision, recall, dan F1-score. Hasil evaluasi model ditunjukkan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Naïve Bayes

Kelas Sentimen	Precision	Recall	F1-Score
Negatif	0,94	0,84	0,89
Netral	0,89	0,98	0,93
Positif	1,00	0,90	0,95

Tabel 3. Metrix dan Hasil Keseluruhan

Metrik Keseluruhan	Hasil
Accuracy	93,00%
Macro Average Precision	0,94
Macro Average Recall	0,91
Macro Average F1-Score	0,92
Weighted Average Precision	0,93
Weighted Average Recall	0,93
Weighted Average F1-Score	0,93

Berdasarkan hasil evaluasi, model Naïve Bayes menghasilkan akurasi sebesar 93,00%. Nilai precision tertinggi diperoleh pada kategori positif (1,00), menunjukkan bahwa seluruh komentar yang diprediksi positif sesuai dengan label sebenarnya. Nilai recall tertinggi terdapat pada kategori netral (0,98), menunjukkan kemampuan model yang sangat baik dalam mengenali komentar netral. Secara keseluruhan, F1-score pada seluruh kategori berada di atas 0,89, menunjukkan keseimbangan yang baik antara precision dan recall.

Untuk menguji konsistensi model, penelitian ini juga menerapkan 5-Fold Cross Validation. Hasil pengujian menunjukkan rata-rata akurasi sebesar 92,00% dengan standar deviasi 1,79%. Nilai standar deviasi yang relatif kecil menunjukkan bahwa model memiliki stabilitas yang baik pada berbagai pembagian data. Dengan demikian, algoritma Naïve Bayes dapat dikatakan mampu mengklasifikasikan sentimen komentar TikTok terkait usaha F&B di Tembilahan secara akurat dan konsisten.

2. Pembahasan

Hasil penelitian menunjukkan bahwa sentimen netral mendominasi komentar pengguna TikTok terhadap usaha F&B di Tembilahan, dengan 248 komentar (49,6%). Dominasi ini menunjukkan bahwa sebagian besar pengguna TikTok cenderung memberikan tanggapan informatif menanyakan harga, lokasi, atau menu atau sekadar berinteraksi tanpa menunjukkan penilaian yang jelas mengarah pada sentimen positif maupun negatif. Fenomena ini menunjukkan bahwa TikTok tidak hanya digunakan sebagai sarana memberikan ulasan produk, tetapi juga sebagai media komunikasi antara konsumen dan pelaku usaha. Tingginya sentimen netral mengindikasikan bahwa aktivitas pemasaran digital yang dilakukan pelaku usaha berhasil menarik perhatian pengguna sehingga memunculkan interaksi yang cukup tinggi pada kolom komentar.

Sentimen positif mencapai 31,2% (156 komentar), umumnya berupa ungkapan kepuasan terhadap rasa makanan dan minuman, kualitas produk, harga yang dianggap sesuai, pelayanan yang baik, serta kenyamanan tempat usaha. Temuan ini menunjukkan bahwa sebagian besar konsumen memiliki persepsi yang baik terhadap usaha F&B yang menjadi objek penelitian, dan

dapat menjadi indikator bahwa strategi pemasaran digital melalui TikTok mampu membangun citra positif serta meningkatkan ketertarikan konsumen. Dalam konteks pemasaran digital, komentar positif berperan penting sebagai bentuk *electronic word of mouth* yang berpotensi memengaruhi keputusan pembelian konsumen lain.

Di sisi lain, sentimen negatif tercatat sebesar 19,2% (96 komentar). Meskipun jumlahnya lebih kecil dibandingkan sentimen positif dan netral, keberadaannya tetap perlu menjadi perhatian pelaku usaha, karena umumnya berkaitan dengan kualitas pelayanan, waktu penyajian, harga, maupun ekspektasi konsumen yang tidak sesuai dengan pengalaman yang diperoleh. Sentimen negatif dapat menjadi sumber informasi berharga bagi pelaku usaha untuk melakukan evaluasi dan perbaikan layanan, termasuk mengidentifikasi aspek yang paling sering dikeluhkan sebagai dasar pengambilan keputusan strategis dalam meningkatkan kualitas produk maupun pelayanan.

Hasil klasifikasi menggunakan Naïve Bayes menunjukkan akurasi sebesar 93,00%, dengan rata-rata akurasi *cross validation* sebesar 92,00%. Nilai ini menunjukkan bahwa model memiliki kemampuan baik dalam mengidentifikasi dan mengelompokkan sentimen konsumen berdasarkan komentar yang dipublikasikan di TikTok. Tingginya nilai *precision*, *recall*, dan *F1-score* pada seluruh kategori sentimen menunjukkan bahwa Naïve Bayes masih relevan digunakan dalam penelitian analisis sentimen, khususnya pada data teks berbahasa Indonesia yang berasal dari media sosial. Hasil ini sejalan dengan penelitian terdahulu yang menyatakan bahwa Naïve Bayes memiliki keunggulan dalam klasifikasi teks karena sederhana, cepat, dan mampu menghasilkan akurasi yang relatif tinggi meskipun menggunakan dataset yang tidak terlalu besar (Layuk et al., 2025; Ghiffary & Syafrullah, 2024; Putra et al., 2025). Sebagai perbandingan, akurasi yang diperoleh pada penelitian ini lebih tinggi dibandingkan penelitian Ghiffary dan Syafrullah (2024), yang melaporkan akurasi model Naïve Bayes sebesar 63,08% dalam menganalisis sentimen pengguna TikTok Shop. Perbedaan ini mengindikasikan bahwa kombinasi *preprocessing* yang lebih menyeluruh dan distribusi data yang digunakan pada penelitian ini turut berkontribusi terhadap performa klasifikasi yang lebih optimal.

Dari perspektif pemasaran digital, hasil penelitian ini memberikan gambaran bahwa analisis sentimen dapat digunakan sebagai alat evaluasi strategi pemasaran yang efektif bagi pelaku usaha F&B di Tembilahan. Informasi mengenai persepsi konsumen yang diperoleh dari media sosial dapat dimanfaatkan untuk memahami kebutuhan pasar, meningkatkan kualitas produk, memperbaiki pelayanan, serta menyusun strategi promosi yang lebih tepat sasaran. Dengan data hasil analisis sentimen, pelaku usaha tidak hanya bergantung pada intuisi dalam mengambil keputusan, tetapi juga dapat menerapkan pendekatan *data-driven decision making* untuk meningkatkan daya saing usaha di era digital.

Secara keseluruhan, penelitian ini menunjukkan bahwa TikTok merupakan sumber data yang potensial untuk menggambarkan persepsi konsumen terhadap usaha F&B. Dominasi sentimen netral dan positif menunjukkan bahwa respons pengguna terhadap usaha F&B di Tembilahan cenderung baik, sedangkan keberadaan sentimen negatif dapat dimanfaatkan sebagai bahan evaluasi untuk perbaikan berkelanjutan. Oleh karena itu, penerapan analisis sentimen menggunakan algoritma Naïve Bayes dapat menjadi alternatif yang efektif dalam mendukung evaluasi dan pengembangan strategi pemasaran digital pada sektor *food and beverage*.

KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Naïve Bayes Classifier untuk menganalisis sentimen konsumen terhadap usaha *Food and Beverage* (F&B) di Tembilahan berdasarkan komentar pada media sosial TikTok. Dari total 500 komentar yang dianalisis, diperoleh distribusi sentimen berupa 248 komentar netral (49,6%), 156 komentar positif (31,2%), dan 96 komentar negatif (19,2%). Hasil ini menunjukkan bahwa mayoritas pengguna TikTok memberikan tanggapan yang bersifat netral, sementara sentimen positif lebih banyak dibandingkan sentimen

negatif mengindikasikan bahwa persepsi konsumen terhadap usaha F&B di Tembilahan cenderung baik, meskipun masih terdapat aspek yang perlu dievaluasi untuk meningkatkan kepuasan pelanggan.

Berdasarkan hasil klasifikasi, algoritma Naïve Bayes mampu menghasilkan akurasi sebesar 93,00%, dengan rata-rata akurasi 5-Fold Cross Validation sebesar 92,00% dan standar deviasi 1,79%. Hasil ini menunjukkan bahwa model memiliki performa yang baik dan stabil dalam mengklasifikasikan sentimen komentar pengguna TikTok ke dalam kategori positif, negatif, dan netral. Nilai precision, recall, dan F1-score yang tinggi pada setiap kategori sentimen juga menunjukkan bahwa metode yang digunakan mampu mengidentifikasi pola sentimen secara efektif.

Secara praktis, hasil penelitian ini menunjukkan bahwa analisis sentimen dapat dimanfaatkan sebagai dasar evaluasi strategi pemasaran digital bagi pelaku usaha F&B di Tembilahan. Informasi yang diperoleh dari komentar konsumen dapat digunakan untuk memahami persepsi pasar, meningkatkan kualitas produk dan layanan, serta menyusun strategi promosi yang lebih sesuai dengan kebutuhan konsumen. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menerapkan pelabelan manual yang lebih komprehensif, serta membandingkan performa Naïve Bayes dengan algoritma lain seperti Support Vector Machine (SVM), Random Forest, atau metode deep learning guna memperoleh hasil klasifikasi yang lebih optimal.

REFERENSI

- Bird, S., Klein, E., & Loper, E. (2021). *Natural Language Processing with Python*. O'Reilly Media.
- Creswell, J. W., & Creswell, J. D. (2023). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.
- DataReportal. (2025). *Digital 2025: Indonesia*. DataReportal (Kepios, We Are Social, Meltwater). <https://datareportal.com/reports/digital-2025-indonesia>
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- Ghiffary, F. A., & Syafrullah, M. (2024). Analisis Sentimen Pengguna TikTok Shop Menggunakan Metode Naive Bayes Classifier. *Jurnal (Cek Nama Jurnal Pada Sumber Asli)*.
- Kurniawan. (2025). Kajian sistematis (systematic literature review) mengenai analisis sentimen media sosial.
- Layuk, (2025). *Penerapan feature selection untuk meningkatkan performa Naive Bayes dalam klasifikasi sentimen*.
- Liu, B. (2022). *Sentiment Analysis and Opinion Mining*. Cambridge University Press.
- Manning, C. D., Raghavan, P., & Schütze, H. (2022). *Introduction to Information Retrieval*. Cambridge University Press.
- Mitchell, T. M. (2021). *Machine Learning*. McGraw-Hill.
- Putra, (2025). Analisis Naive Bayes terhadap persepsi masyarakat pada fenomena bisnis digital di media sosial.
- Sekaran, U., & Bougie, R. (2020). *Research Methods for Business: A Skill-Building Approach*. Wiley.
- Social, W. A., & Meltwater. (2024). *Digital 2024: Global Overview Report*. We Are Social & Meltwater. <https://wearesocial.com/us/blog/2024/01/digital-2024/>
- Taherdoost, H. (2020). Sampling methods in research methodology; how to choose a sampling technique for research. *International Journal of Academic Research in Management (IJARM)*, 5(2), 18–27.
- Tala, F. Z. (2021). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Universiteit van Amsterdam.
- Tsiourlini, M., Tzafilkou, K., Karapiperis, D., & Tjortjis, C. (2024). Text Analytics on YouTube Comments for Food Products. *Information*, 15(6).
- Visalli, A. (2026). Sentiment analysis in consumer science: An emerging approach to understanding affective meaning in digital opinions.