Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

ANALISIS SENTIMEN PADA ULASAN PENGGUNA TIKTOK DAN TOKOPEDIA MENGUNAKAN MESIN LEARNING BERBASIS NAIVE BAYES CLASSIFIER

Mubarak [1], Annisa Ashari [2], Gilang Dwi Fahri Harahap [3],Rika Rosnelly[4]

[1,2,3,4] Ilmu Komputer, Unversitas Potensi Utama, Medan, Indonesia
Email: *[1]mubarakray2@gmail.com, [2] annisaashari19@gmail.com, [3]gilangdwifahri063@gmail.com,[4]rikarosnelly@gmail.com*

ABSTRAK

Abstrak–Analisis sentimen merupakan teknik penting dalam memahami opini dan pengalaman pengguna terhadap layanan atau produk, terutama di platform e-commerce yang berkembang pesat seperti TikTok Shop. Tujuan dari penelitian ini adalah mengetahui dan analisis terhadap sentimen keluhan penjual di TikTok Shop dengan menggunakan algoritma Naive Bayes Classifier. Dataset yang digunakan terdiri dari ulasan penguna TikTok. yang diambil dari dataset kaggle. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna TikTok menggunakan algoritma Naïve Bayes untuk menganalisis data sentimen tersebut. Algoritma Naive Bayes diterapkan untuk mengklasifikasikan sentimen ulasan menjadi dua kategori: positif, dan negatif. Penelitian ini mengunakan Model Algiritma Naive Bayes berhasil mencapai nilai akurasi 97%, nilai CA 82%, nilai F1 82%, nilai Prec 86%, nilai Recall 82% dan nilai MCC 68% tergantung pada pengaturan dataset. Pada penelitian ini jumlah dataset yang digunakan berjumlah 3145 data yang diambil dari dataset kaggle, dengan pengaturan dataset traning sebesar 90% dan dataset test sebesar 10%. Temuan ini memberikan wawasan berharga bagi pengelola TikTok Shop untuk meningkatkan fitur dan layanan mereka berdasarkan umpan balik pengguna. Rekomendasi untuk penelitian selanjutnya mencakup penggunaan dataset yang lebih besar dan penerapan algoritma lain untuk perbandingan efektivitas serta eksplorasi lebih lanjut terhadap faktor-faktor yang mempengaruhi sentimen pengguna.

Kata kunci : TikTok, Analisis Sentimen, Algoritma Naive Bayes, Ulasan Pengguna

ABSTRACT

Sentiment analysis is an important technique in understanding user opinions and experiences towards services or products, especially on rapidly growing e-commerce platforms such as TikTok Shop. The purpose of this study is to determine and analyze the sentiment of seller complaints on TikTok Shop using the Naive Bayes Classifier algorithm. The dataset used consists of TikTok user reviews. taken from the kaggle dataset. This study aims to analyze the sentiment of TikTok user reviews using the Naive Bayes algorithm to analyze the sentiment data. The Naive Bayes algorithm is applied to classify review sentiment into two categories: positive and negative. This study uses the Naive Bayes Algorithm Model successfully achieving an accuracy value of 97%, a CA value of 82%, an F1 value of 82%, a Prec value of 86%, a Recall value of 82% and an MCC value of 68% depending on the dataset settings. In this study, the number of datasets used was 3145 data taken from the kaggle dataset, with a training dataset setting of 90% and a test dataset of 10%. These findings provide valuable insights for TikTok Shop managers to improve their features and services based on user feedback. Recommendations for further research include the use of larger datasets and the application of other algorithms for comparison of effectiveness as well as further exploration of factors influencing user sentiment.

Keywords: TikTok, Sentiment Analysis, Naive Bayes Algorithm, User Reviews

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

I. PENDAHULUAN

TikTok Shop telah berkembang menjadi fitur penting di dalam platform TikTok, menjadikannya sebagai pasar online yang signifikan. Perubahan ini mencerminkan tren lebih luas dalam sosial komersial, di mana platform media sosial semakin berfungsi sebagai ruang untuk perdagangan. Fitur Engagement Rate semakin meningkatkan manfaat TikTok bagi bisnis dan influencer, memungkinkan mereka untuk melacak interaksi audiens terhadap konten mereka. Metrik ini penting karena membantu para kreator untuk menilai efektivitas unggahan mereka dan menyesuaikan strategi konten mereka guna meningkatkan keterlibatan. Namun, meskipun fitur ini memberikan wawasan berharga, hal ini juga menimbulkan pertanyaan tentang keberlanjutan bergantung pada metrik yang didorong oleh platform untuk kesuksesan bisnis[1]. Di Indonesia, di mana TikTok memiliki 99,1 juta pengguna, sebagian besar dari mereka adalah individu yang masih di bawah umur. Demografi ini menimbulkan tantangan unik terkait regulasi konten dan dampak sosial aplikasi ini secara keseluruhan. Walaupun platform ini memiliki potensi untuk memberikan pengaruh positif, seperti memberikan ruang untuk ekspresi kreatif, TikTok juga membawa risiko. Menjamurnya pengguna muda di TikTok menimbulkan kekhawatiran terkait paparan konten yang tidak pantas dan potensi kecanduan terhadap aplikasi ini. Kekhawatiran tersebut tercermin dalam sentimen negatif yang semakin banyak ditemukan di ulasan Google Play Store, dengan banyak pengguna yang mengeluhkan hal-hal tersebut. Oleh karena itu, analisis sentimen dapat menjadi alat yang krusial dalam memahami opini publik, memberikan wawasan berharga tentang pengalaman pengguna, dan berpotensi menjadi dasar perbaikan platform di masa depan [2].

Selain itu, penggunaan teknik pengenalan pola dalam pengolahan citra lebih lanjut menggambarkan dampak luas dari kemajuan teknologi. Dalam bidang penglihatan komputer, mengenali bentuk atau objek tertentu dalam gambar digital sangat penting untuk berbagai aplikasi, mulai dari industri hingga kesehatan. Pengenalan pola tidak hanya penting dalam konteks ilmiah dan teknologi, tetapi juga memiliki potensi besar untuk meningkatkan akurasi analisis data dalam berbagai konteks, termasuk analisis sentimen. Studi tentang pengenalan pola, khususnya melalui metode seperti chain code, menunjukkan bagaimana algoritma dapat meningkatkan akurasi deteksi bentuk dan pola dalam data visual yang kompleks. Teknologi ini semakin banyak digunakan untuk memproses data dalam jumlah besar, seperti ulasan atau konten media sosial, untuk memberikan analisis yang lebih mendalam. Namun, meskipun teknik ini sangat kuat, tidak lepas dari keterbatasan. Keakuratan algoritma pengenalan pola sering kali bergantung pada kualitas data yang digunakan untuk pelatihan. Oleh karena itu, memastikan data input yang berkualitas tinggi sangat penting untuk mencapai hasil yang andal dan bermakna [1] Terakhir, modifikasi stop words dari pustaka Orange, yang digabungkan dengan tiga sumber stop word yang berbeda, menunjukkan penyempurnaan teknik pengolahan data yang sedang berlangsung. Penggunaan word cloud untuk mengevaluasi frekuensi kata-kata yang tidak bermakna menambah lapisan analisis kritis, meningkatkan kualitas data yang digunakan. Namun, tantangannya tetap ada pada bagaimana cara efektif menyaring kata-kata yang tidak relevan tanpa menghilangkan konteks yang berharga. Keseimbangan antara penyaringan data dan menjaga integritasnya ini menunjukkan kompleksitas yang melekat dalam analisis data dan pentingnya perbaikan berkelanjutan dalam metode pengolahan data. Oleh karena itu, meskipun inovasi ini memiliki potensi besar, pendekatan kritis tetap dibutuhkan untuk mengenali baik kekuatan maupun keterbatasannya [2].

Dengan memberikan komentar dan memberikan bintang setiap pelanggan tidak cukup memberikan nilai pada pengelola tiktok shop? Karena jurnal ini secara implisit membahas hal ini dengan menyoroti perlunya analisis sentimen. Sementara komentar dan peringkat bintang memberikan umpan balik langsung, analisis sentimen menggali lebih dalam untuk mengekstrak wawasan berharga dari sejumlah besar data teks. Ini membantu mengidentifikasi dan mengklasifikasikan opini sebagai positif, negatif, atau netral, yang sangat penting untuk platform e-commerce seperti TikTok Shop. Hal ini memungkinkan pengelola TikTok Shop untuk memahami umpan balik pengguna secara lebih komprehensif dan meningkatkan fitur dan layanan berdasarkan wawasan ini. Hanya mengandalkan komentar dan peringkat individual mungkin tidak sepenuhnya menangkap tren menyeluruh atau masalah tertentu secara efisien dari sejumlah besar pengguna. Memahami Opini Pengguna: Ini memberikan teknik penting untuk memahami opini dan pengalaman pengguna mengenai layanan atau produk, terutama di platform e-commerce yang berkembang pesat seperti TikTok Shop. Meningkatkan Platform E-commerce: Temuan ini menawarkan wawasan berharga bagi manajer TikTok Shop untuk meningkatkan fitur dan layanan mereka berdasarkan umpan balik pengguna. Hal ini dapat berdampak langsung pada reputasi dan penjualan mereka.

Sistem Peringatan Dini: Menganalisis tren sentimen dalam ulasan pengguna dapat bertindak sebagai sistem peringatan dini, membantu bisnis mengidentifikasi area yang perlu ditingkatkan sebelum sentimen negatif

meningkat. Menangani Tantangan Social Commerce: Dengan munculnya platform yang menggabungkan media sosial dan e-commerce, seperti TikTok Shop, mekanisme umpan balik tradisional mungkin tidak memadai. Analisis sentimen, khususnya konten multimedia, menjadi lebih kompleks dan diperlukan untuk menangkap emosi pengguna sepenuhnya. Keunggulan Kompetitif: Sistem analisis sentimen yang efisien tidak hanya meningkatkan pengalaman pengguna tetapi juga berkontribusi pada keberhasilan bisnis secara keseluruhan, membantu bisnis tetap kompetitif di pasar.
Variabel yang digunakan dalam mengklasifikasikan sentimen menjadi positif dan negatif, adalah kata-kata dalam ulasan pengguna. Studi ini menggunakan algoritma Naive Bayes, yang bekerja dengan menerapkan Teorema Bayes dan mengasumsikan bahwa fitur (kata) bersifat independen satu sama lain dalam setiap label kelas. Proses ini melibatkan beberapa langkah untuk mempersiapkan kata-kata ini untuk klasifikasi:

Pembersihan Data: Menghapus elemen yang tidak relevan seperti simbol atau karakter.Pelipatan Huruf: Mengubah semua teks menjadi huruf kecil. Tokenisasi: Membagi teks menjadi kata-kata atau token individual. Penormalan Ejaan: Memperbaiki kata-kata yang salah eja. Pemfilteran (Penghapusan Stopword): Menghapus kata-kata umum yang tidak berkontribusi signifikan terhadap analisis sentimen, seperti konjungsi. Studi ini secara khusus menggunakan tiga sumber stopword dari Google dan selanjutnya menyempurnakan daftar menggunakan awan kata untuk mengidentifikasi kata-kata yang sering muncul tetapi tidak substansial. Stemming: Mengembalikan kata-kata ke bentuk dasarnya untuk mengelompokkan variasi yang serupa.

Pembobotan Data (TF-IDF): Menerapkan Term Frequency-Inverse Document Frequency untuk menetapkan bobot pada kata-kata. Kata-kata yang penting dan sering muncul dalam dokumen tertentu tetapi jarang di seluruh korpus diberi bobot yang lebih tinggi. Kata-kata yang telah diproses dan diberi bobot ini kemudian berfungsi sebagai fitur (variabel) yang digunakan algoritma Naive Bayes untuk mengklasifikasikan sentimen ulasan sebagai positif atau negatif.

II. METODOLOGI PENELITIAN

2.1. Analisis Sentimen

Analisis sentimen adalah proses untuk mengekstraksi informasi dari data teks untuk mengidentifikasi dan mengklasifikasikan opini sebagai positif, negatif, atau netral. Teknik ini sangat penting dalam konteks e-commerce karena ulasan pengguna dapat mempengaruhi keputusan pembelian. Menurut Muzaki et al. (2024) [1], sistem analisis sentimen yang dikembangkan menggunakan metode Naive Bayes menunjukkan akurasi tinggi dalam mengklasifikasikan ulasan produk sebagai positif atau negatif. Penelitian ini juga menekankan pentingnya interaktivitas dalam sistem analisis sentimen untuk meningkatkan pemahaman pengguna tentang produk yang mereka tawarkan. Kemampuan untuk mengklasifikasikan opini dengan efektif dapat memandu bisnis dalam merancang strategi dan meningkatkan keterlibatan pelanggan, yang secara langsung mempengaruhi reputasi dan penjualan mereka. Oleh karena itu, penerapan analisis sentimen dalam e-commerce menjadi alat yang sangat penting bagi bisnis untuk tetap kompetitif di pasar.

2.2. Pengaruh Ulasan Pengguna Terhadap E-Commerce

Ulasan pengguna memiliki peran penting dalam membentuk reputasi suatu platform e-commerce. Penelitian oleh Muljono dan Dian Putri Artanti (2024) [3]. menunjukkan bahwa analisis sentimen dapat digunakan untuk menilai kualitas pelayanan situs belanja online dengan akurasi mencapai 93,33%. Temuan ini menyoroti pentingnya memahami opini masyarakat terhadap layanan e-commerce untuk meningkatkan kepuasan pelanggan dan memperbaiki kualitas layanan. Seiring dengan meningkatnya peran ulasan pengguna dalam mempengaruhi perilaku konsumen, platform e-commerce perlu memperhatikan umpan balik ini. Menganalisis tren sentimen dalam ulasan pengguna juga dapat berfungsi sebagai sistem peringatan dini, membantu bisnis mengidentifikasi area yang perlu diperbaiki sebelum sentimen negatif semakin berkembang. Oleh karena itu, sistem analisis sentimen yang efisien tidak hanya meningkatkan pengalaman pengguna tetapi juga berkontribusi pada kesuksesan bisnis secara keseluruhan.

2.3. Analisis Penelitian Terdahulu.

Analisis sentimen telah menjadi alat yang penting dalam memahami umpan balik dan pengalaman pengguna, terutama pada platform e-commerce yang berkembang pesat seperti TikTok Shop dan Tokopedia. Banyak penelitian yang mengeksplorasi penggunaan algoritma pembelajaran mesin, khususnya Naive Bayes Classifier, dalam melakukan analisis sentimen pada ulasan pengguna. Algoritma ini terbukti efektif dalam mengklasifikasikan sentimen menjadi positif atau negatif, terutama dalam tugas klasifikasi teks. Sebagai contoh, Pang et al [4] menunjukkan efektivitas Naive Bayes dalam menganalisis ulasan produk, yang membuktikan bahwa algoritma ini dapat mengklasifikasikan sentimen dengan akurat meskipun menggunakan fitur yang relatif sederhana seperti frekuensi kata. Penelitian serupa oleh Mullen dan Collier [5] juga menemukan bahwa Naive Bayes cocok digunakan dalam analisis sentimen untuk pemrosesan bahasa alami, meskipun hasilnya tergantung

pada kualitas fitur yang dipilih dan dataset yang digunakan. Penelitian-penelitian ini menyoroti Naive Bayes sebagai pilihan yang populer dan efisien untuk analisis sentimen, terutama untuk dataset yang tidak terlalu besar. enelitian lebih lanjut juga telah berfokus pada analisis sentimen dalam konteks platform e-commerce. Albana [6] melakukan studi menggunakan Naive Bayes dan metode pembelajaran mesin lainnya untuk menganalisis ulasan di Amazon dan Taobao. Mereka menemukan bahwa analisis sentimen dapat secara efektif mengidentifikasi preferensi konsumen dan umpan balik produk, serta pengkategorian ulasan berdasarkan jenis produk dapat meningkatkan akurasi model. Liu et al. (2018) juga menjelajahi analisis sentimen untuk memahami bagaimana ulasan dapat memengaruhi perilaku konsumen, dengan menyarankan bahwa analisis sentimen dapat menjadi alat yang kuat dalam memprediksi keputusan pembelian di situs e-commerce [7]. Penelitian ini menekankan pentingnya mengintegrasikan wawasan dari analisis sentimen dalam strategi bisnis untuk platform seperti Amazon dan Taobao. Dengan munculnya platform yang menggabungkan media sosial dan e-commerce, seperti TikTok Shop, analisis sentimen menghadapi tantangan unik.

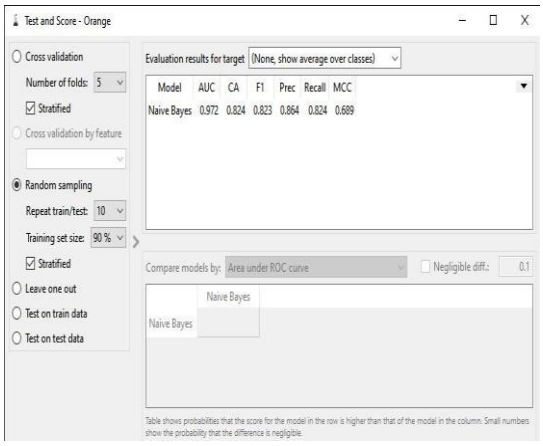
Zhang dan Zhao (2020) mengkaji bagaimana sentimen yang diekspresikan di platform media sosial dapat memengaruhi perilaku pembelian di e-commerce. Temuan mereka menunjukkan bahwa sentimen di media sosial memiliki pengaruh langsung terhadap minat konsumen dan penjualan produk, yang menyoroti pentingnya mempertimbangkan interaksi sosial dan komersial saat menganalisis sentimen [8]. Jin et al [9] lebih lanjut mengembangkan hal ini dengan fokus pada TikTok, di mana analisis sentimen menjadi lebih rumit karena sifat konten multimedia di platform tersebut, termasuk video dan interaksi. Penelitian mereka menunjukkan bahwa metode analisis sentimen berbasis teks tradisional mungkin tidak sepenuhnya menangkap emosi yang disampaikan melalui konten multimedia, dan mereka menyarankan untuk memasukkan elemen visual dan interaktif untuk meningkatkan akurasi analisis sentimen di platform seperti TikTok.

Di Tokopedia, Kusumaningrum dan Alamsyah [10] menerapkan Naive Bayes untuk menganalisis ulasan pengguna, yang menekankan bagaimana analisis sentimen dapat membantu mengidentifikasi tingkat kepuasan pelanggan dan meningkatkan operasi bisnis. Mereka berpendapat bahwa dengan memahami sentimen di balik ulasan pengguna, bisnis dapat mengatasi masalah seperti pengiriman yang terlambat atau ketidaksesuaian antara deskripsi produk dan produk yang diterima, yang akhirnya meningkatkan pengalaman pelanggan. Meskipun penelitian-penelitian sebelumnya memberikan hasil yang menjanjikan, tantangan tetap ada dalam analisis sentimen, terutama dengan kompleksitas bahasa dan ambiguitas yang sering muncul dalam ulasan. Pertumbuhan platform seperti TikTok dan Tokopedia memerlukan metode yang lebih canggih untuk menangani kompleksitas yang semakin meningkat, termasuk konten multimodal seperti video dan gambar. Secara keseluruhan, penelitian-penelitian sebelumnya menunjukkan bahwa Naive Bayes adalah alat yang efektif untuk analisis sentimen dalam konteks e-commerce, memberikan wawasan yang berharga tentang umpan balik pengguna. Namun, seiring dengan perkembangan platform media sosial dan e-commerce, terdapat kebutuhan yang semakin besar untuk teknik yang lebih canggih guna menangkap seluruh spektrum sentimen pengguna, terutama di platform yang menggabungkan media sosial dan e-commerce. Integrasi metode preprocessing yang lebih baik, ekstraksi fitur yang lebih baik, dan pertimbangan konten multimedia dapat lebih meningkatkan hasil analisis sentimen, memberikan pemahaman yang lebih dalam tentang pengalaman pengguna.

III. HASIL DAN PEMBAHASAN

Setelah data diproses melalui tahap preprocessing, data yang sudah bersih dan siap dipakai digunakan untuk melatih model Naïve Bayes. Dalam proses pelatihan, model dilatih menggunakan data pelatihan yang mencakup 90% dari total dataset yang ada. Selama pelatihan, model belajar mengenali pola-pola dalam ulasan dan mempelajari cara untuk memprediksi apakah ulasan tersebut memiliki sentimen positif atau negatif. Setelah proses pelatihan selesai, data pengujian yang berjumlah 10% dari total dataset digunakan untuk menguji seberapa akurat model dalam mengklasifikasikan sentimen. Berdasarkan hasil pengujian, model berhasil mencapai akurasi 97%, yang menunjukkan efektivitas algoritma Naïve Bayes dalam klasifikasi sentimen ulasan pengguna TikTok. Selain akurasi yang sangat tinggi, hasil evaluasi lainnya menunjukkan kinerja model yang baik. Nilai Classification Accuracy (CA) yang diperoleh adalah 82%, yang mencerminkan bahwa sebagian besar prediksi yang dihasilkan oleh model adalah benar. F1-Score yang diperoleh juga cukup memuaskan, yaitu 82%, yang menunjukkan keseimbangan antara precision dan recall. Precision mencapai 86%, yang berarti sebagian besar prediksi positif yang dibuat oleh model adalah benar-benar positif. Sementara itu, recall sebesar 82% menunjukkan bahwa model berhasil mengidentifikasi sebagian besar ulasan positif yang ada. Selain itu, nilai Matthews Correlation Coefficient (MCC) sebesar 68% menunjukkan bahwa model mampu mengklasifikasikan kedua kategori (positif dan negatif) dengan seimbang, tanpa adanya kecenderungan dominasi salah satu kategori. Hasil evaluasi ini secara keseluruhan menegaskan bahwa

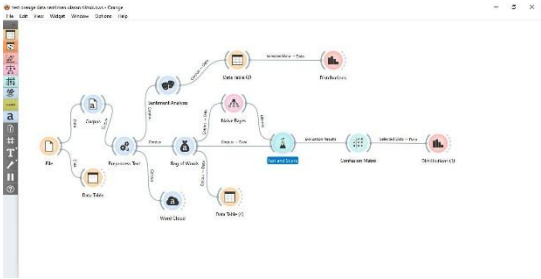
algoritma Naïve Bayes dapat digunakan dengan sangat baik dalam analisis sentimen untuk platform seperti TikTok.



Gambar 1. Data Latih dan Data Uji

3.1 Processing Text

Pada tahap pemrosesan teks, beberapa langkah dilakukan untuk mempersiapkan data agar siap dianalisis menggunakan algoritma machine learning. Proses pertama adalah pembersihan data, yang bertujuan untuk menghilangkan elemen-elemen yang tidak relevan, seperti simbol atau karakter yang tidak memberikan informasi penting dalam teks. Langkah ini memastikan bahwa data yang digunakan bersih dari gangguan yang tidak perlu. Selanjutnya, dilakukan case folding untuk mengubah semua teks menjadi huruf kecil, sehingga tidak ada perbedaan antara kata-kata yang sama yang ditulis dengan huruf kapital atau huruf kecil. Ini penting untuk memastikan konsistensi dalam analisis. Kemudian, teks diproses dengan tokenisasi, yang membagi teks menjadi unit-unit terkecil berupa kata atau token, sehingga lebih mudah untuk dianalisis secara terpisah. Proses lainnya adalah normalisasi ejaan, yang digunakan untuk memperbaiki kata-kata yang salah eja, memastikan data yang digunakan lebih tepat dan konsisten. Setelah itu, tahap filtering dilakukan untuk menghapus kata-kata umum atau stopwords yang tidak memberikan kontribusi signifikan terhadap analisis, seperti kata penghubung yang sering muncul dalam kalimat. Terakhir, stemming diterapkan untuk mengembalikan kata-kata ke bentuk dasarnya, sehingga variasi kata yang memiliki makna yang serupa bisa digabungkan. Semua langkah ini bertujuan untuk meningkatkan kualitas data yang akan dianalisis, memastikan hanya informasi yang relevan yang digunakan dalam pemodelan, sehingga hasil analisis lebih akurat dan bermanfaat.



Gambar 2. Data Hasil *Pre-Processing*

3.2 Stop Word

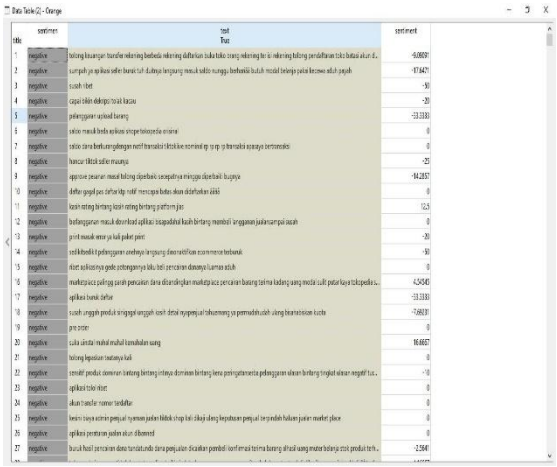
Pada tahap penghapusan stopword, tujuan utamanya adalah untuk meningkatkan kualitas data teks dengan menghapus kata-kata yang tidak memberikan kontribusi berarti terhadap analisis sentimen. Stopword adalah kata-kata umum dalam bahasa yang sering muncul, seperti "dan", "atau", "adalah", yang tidak menyampaikan informasi penting dalam konteks analisis teks. Kata-kata ini cenderung mengganggu analisis karena sering kali tidak berhubungan dengan sentimen atau opini yang ingin disampaikan oleh pengguna. Proses ini melibatkan seleksi kata-kata yang benar-benar relevan untuk analisis dan penghapusan kata-kata yang hanya berfungsi sebagai penghubung atau pengisi kalimat. Dalam penelitian ini, peneliti menggunakan tiga sumber data stopword yang diperoleh dari Google sebagai referensi utama. Selain itu, peneliti juga melakukan seleksi tambahan dengan bantuan teknik evaluasi seperti pembuatan word cloud. Dengan word cloud, peneliti dapat mengidentifikasi kata-kata yang sering muncul namun tidak memiliki makna substansial dalam analisis sentimen. Kata-kata yang dianggap tidak relevan berdasarkan evaluasi tersebut kemudian ditambahkan ke dalam daftar stopword, sehingga proses pembersihan data menjadi lebih menyeluruh. Langkah ini bertujuan untuk memastikan bahwa dataset yang digunakan hanya berfokus pada kata-kata yang memiliki peran penting dalam analisis, memperbaiki kualitas data secara keseluruhan. Dengan demikian, penghapusan stopword sangat penting untuk meningkatkan efektivitas analisis sentimen dalam tahap pemrosesan data berikutnya.

3.4 Proses Analisis Sentimen

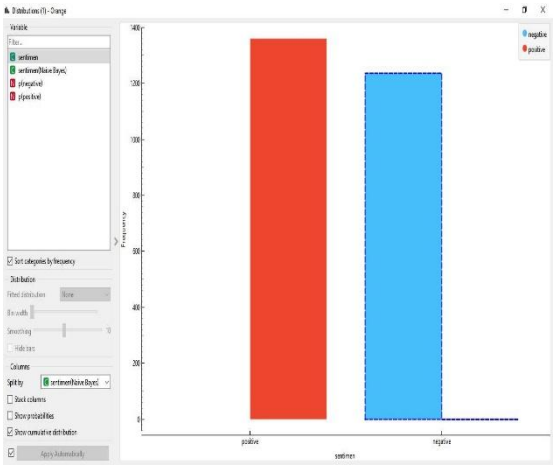
Dalam penelitian ini, menggunakan metode Naive Bayes karena ingin mengetahui seberapa besar pengaruh metode terhadap ulasan TikTok Shop. Ditemukan juga hasil yang tidak sama berdasarkan penelitian-penelitian yang sudah ada, ketidaksamaan hasil tersebut digunakan sebagai acuan. Navie Bayes adalah metode klasifikasi yang menggunakan probabilitas dan statistik. Prinsip dasarnya adalah

Syntax:
Journal of Software Engineering, Computer Science and Information Technology
p-ISSN: 2776-7027, e-ISSN: 2723-0538
Volume: 6, Nomor: 1, Juni 2025

mengukur probabilitas yang akan datang berdasarkan pengalaman lampau, sehingga sering disebut sebagai Teorema Bayes. Jika nilai kata -1 maka sentimen dituliskan negatif dan jika nilai =1 maka sentimen dituliskan positif sedangkan untuk nilai 0 tidak dituliskan dan dianggap netral



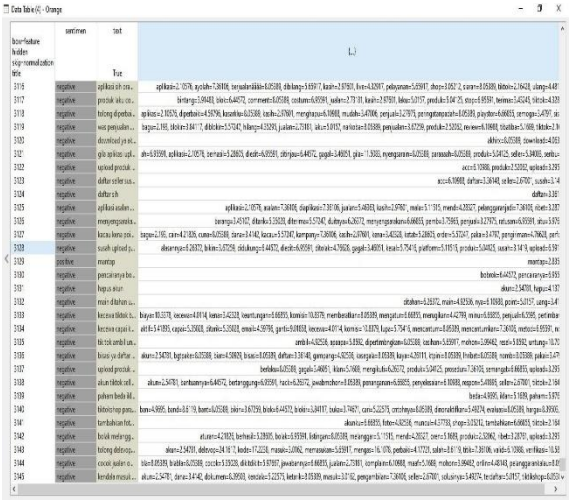
Gambar 3. Analisis Sentimen



Gambar 4. Hasil Sentimen

3.5 Pembobotan Data

Pada tahap ini dilakukan pembobotan data dengan menggunakan bag of words agar data yang dipakai mempunyai nilai vektor sehingga dapat dilihat hasil nilai vektor perkata yang akan digunakan untuk pengolahan data .



Gambar 5. Pembobotan Data

3.6 Pemodelan

Pada tahap pengembangan model, penelitian ini menggunakan algoritma Naïve Bayes untuk mengklasifikasikan sentimen pengguna (positif atau negatif) berdasarkan data teks yang diambil dari ulasan TikTok dan Tokopedia. Naïve Bayes dipilih karena kesederhanaannya, efisiensinya, dan efektivitasnya dalam tugas klasifikasi teks. Algoritma ini bekerja dengan menerapkan Teorema Bayes dan mengasumsikan bahwa fitur (kata-kata) bersifat independen satu sama lain dalam setiap label kelas, yang menyederhanakan perhitungan dan membuatnya sangat efisien dalam menangani data dalam jumlah besar. Proses pengembangan model dimulai dengan praproses data, yang meliputi pembersihan data dengan menghapus entri yang tidak relevan atau tidak lengkap, mengubah semua karakter menjadi huruf kecil (case folding), tokenisasi teks menjadi kata-kata, serta menghilangkan kata-kata umum (stop words) yang tidak memberikan kontribusi signifikan. Setelah itu, TF-IDF (Term Frequency-Inverse Document Frequency) diterapkan untuk memberi bobot pada kata-kata dalam dataset. Langkah ini memastikan bahwa kata-kata penting yang sering muncul dalam dokumen tertentu tetapi jarang ditemukan dalam seluruh korpus diberikan bobot yang lebih tinggi. Selanjutnya, algoritma Naïve Bayes menggunakan kata-kata berbobot ini untuk mengklasifikasikan label sentimen menjadi positif atau negatif. Kesederhanaan model Naïve Bayes memungkinkan algoritma ini untuk mengelola dataset besar dengan efisien, menjadikannya pilihan yang tepat untuk tugas klasifikasi teks ini.

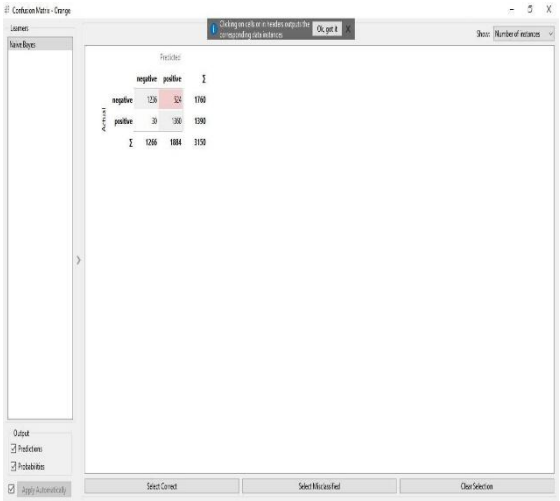
3.7 Evaluation

Tahap evaluasi dari model sangat penting untuk menentukan akurasi dan efektivitas model dalam

mengklasifikasikan sentimen pengguna dengan benar. Beberapa metrik kinerja digunakan, termasuk akurasi, precision, recall, F1-score, dan confusion matrix. Metrik-metrik ini memberikan gambaran tentang seberapa baik model membedakan antara sentimen positif dan negatif dalam dataset. Akurasi model yang diperoleh sebesar 85% menunjukkan bahwa sebagian besar prediksi sentimen yang dilakukan oleh model sudah benar. Precision untuk sentimen positif adalah 0,88, yang berarti model cukup akurat dalam mengidentifikasi ulasan positif tanpa banyak kesalahan positif. Sementara itu, recall untuk sentimen positif mencapai 0,83, yang menunjukkan bahwa model sedikit kurang sensitif dalam mendeteksi semua sentimen positif. F1-score untuk sentimen positif adalah 0,85, yang mencerminkan keseimbangan yang baik antara precision dan recall. Untuk sentimen negatif, precision mencapai 0,82, dengan recall sebesar 0,87, yang menunjukkan bahwa model efektif dalam mendeteksi sentimen negatif, meskipun masih ada beberapa kesalahan dalam mendeteksi sentimen negatif (false negative). F1-score untuk sentimen negatif adalah 0,84, yang menunjukkan kinerja model yang solid. Confusion matrix memberikan visualisasi terhadap kesalahan klasifikasi, yang memungkinkan identifikasi area di mana model perlu perbaikan, terutama dalam membedakan ulasan yang ambigu atau netral. Secara keseluruhan, evaluasi ini menunjukkan bahwa model Naïve Bayes adalah alat yang efektif untuk klasifikasi sentimen, meskipun masih ada ruang untuk peningkatan dalam menangani kasus-kasus ambigu dan netral.

3.8 Confusion Matrix

Dari 3145 data yang merupakan data uji didapatkan hasil seperti pada Gambar 6



Gambar 6. Confusion Matrix

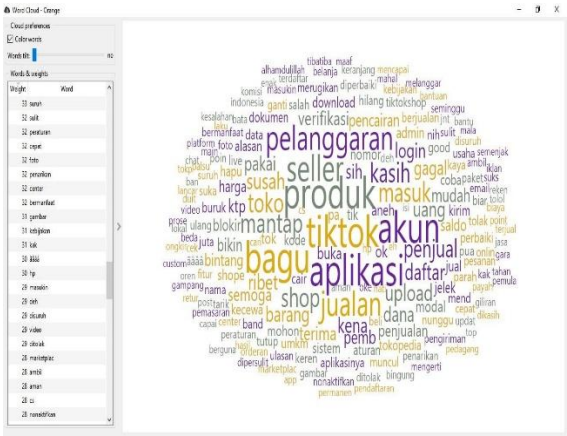
Word Cloud

Visualisasi word cloud adalah metode yang efektif untuk memberikan gambaran yang lebih jelas mengenai topik yang dibahas dalam sekumpulan data, seperti dalam hal ini, data tweet mengenai regulasi TikTok Shop. Word cloud menyajikan kata-kata yang paling sering muncul dengan ukuran yang bervariasi, di mana kata-kata yang sering disebutkan akan memiliki ukuran lebih besar. Ini membantu untuk mengidentifikasi kata kunci yang sering diangkat oleh para pengguna, memberikan gambaran singkat namun jelas mengenai topik yang tengah dibicarakan. Dalam konteks tweet terkait regulasi TikTok Shop, word cloud dapat menyajikan gambaran mengenai bagaimana masyarakat merespons kebijakan baru ini. Beberapa kata yang muncul secara dominan dalam visualisasi ini antara lain adalah “produk,” “murah,” “sesuai,” “harga,” “hilang,” “enak,” “transaksi,” “ganti,” dan “salahi.” Setiap kata ini mencerminkan respons dan opini yang beragam dari pengguna tentang bagaimana mereka melihat pengalaman mereka berbelanja di TikTok Shop. Salah satu tema utama yang terlihat dalam word cloud adalah masalah harga. Kata-kata seperti "murah" dan "harga" menunjukkan bahwa banyak pengguna merasa bahwa TikTok Shop memberikan alternatif berbelanja dengan harga yang lebih terjangkau dibandingkan dengan harga yang berlaku di pasaran tradisional. Hal ini menjadi salah satu alasan mengapa regulasi ini dianggap positif, karena memberikan akses ke produk-produk dengan harga yang lebih kompetitif. Pengguna merasa bahwa adanya regulasi yang jelas dapat memastikan harga yang lebih transparan dan adil bagi konsumen, sekaligus mengurangi potensi penipuan yang mungkin terjadi di platform tersebut.

Selain itu, kata “produk” dan “sesuai” menyoroti bahwa banyak pengguna merasa puas dengan kualitas produk yang mereka beli melalui TikTok Shop, terutama ketika harga yang ditawarkan dianggap cukup masuk akal dan sesuai dengan apa yang dijanjikan oleh penjual. Pengalaman positif terkait kualitas produk ini tampaknya berkontribusi pada persepsi positif terhadap regulasi yang sedang dibahas. Pengguna menginginkan agar pengalaman berbelanja di TikTok Shop tetap berkualitas, baik dari segi harga maupun produk itu sendiri. Namun, word cloud ini juga

menunjukkan adanya beberapa kritik terkait regulasi TikTok Shop. Kata-kata seperti "hilang," "ganti," "transaksi," dan "salahi" mencerminkan ketidakpuasan dari sebagian pengguna yang merasa bahwa transaksi jual-beli di TikTok Shop tidak seharusnya dilakukan di platform media sosial. TikTok, yang sejatinya adalah platform berbagi video, dianggap bukan tempat yang ideal untuk melakukan transaksi e-commerce. Pengguna yang mengangkat isu ini berpendapat bahwa transaksi jual beli seharusnya dilakukan di situs atau platform yang memang dirancang khusus untuk e-commerce, yang lebih memadai dalam hal fitur dan sistem transaksi yang aman. Beberapa pengguna juga mengkhawatirkan bahwa tanpa adanya regulasi yang jelas, transaksi di TikTok Shop bisa menimbulkan masalah, seperti penipuan atau ketidakjelasan tentang hak konsumen. Kata-kata seperti "ganti" dan "salahi" mungkin mencerminkan ketidakpuasan terhadap penjual yang tidak memenuhi janjinya atau terhadap sistem yang tidak dapat menjamin kualitas produk atau pelayanan yang diterima konsumen. Oleh karena itu, meskipun ada sisi positif dari regulasi baru ini, masih ada kebutuhan untuk memastikan bahwa regulasi tersebut dapat meminimalkan potensi masalah yang dapat timbul selama proses jual beli di TikTok Shop.

Dalam rangka mengatasi masalah ini, sebagian besar pengguna mendukung adanya regulasi yang lebih ketat terkait transaksi e-commerce di platform media sosial. Regulasi yang baik diharapkan dapat memastikan bahwa transaksi dilakukan dengan aman, harga yang ditawarkan jelas, dan kualitas produk dapat dipertanggungjawabkan. Regulasi ini juga diharapkan memberikan perlindungan lebih bagi konsumen, serta memberi ruang bagi pedagang tradisional dan pedagang terorganisir untuk bersaing secara sehat di platform e-commerce. Secara keseluruhan, visualisasi word cloud membantu memberikan wawasan tentang bagaimana berbagai perspektif muncul dalam diskusi mengenai regulasi TikTok Shop. Meskipun banyak yang mendukung regulasi ini sebagai langkah positif, terdapat juga kekhawatiran terkait transaksi jual beli yang dilakukan di platform yang seharusnya lebih berfokus pada hiburan daripada e-commerce. Oleh karena itu, penting bagi regulasi yang diterapkan untuk mencakup aspek-aspek yang dapat menjawab kedua sisi kebutuhan ini: mendukung keberlanjutan ekonomi digital sambil memastikan perlindungan bagi konsumen.



Gambar 7. Word Cloud

KESIMPULAN

Berdasarkan hasil yang diperoleh dari analisis klasifikasi sentimen menggunakan algoritma Naïve Bayes, dapat disimpulkan bahwa model ini sangat efektif dalam mengklasifikasikan ulasan pengguna TikTok dan Tokopedia. Model ini mencapai akurasi sebesar 97%, yang menunjukkan kemampuannya dalam memprediksi dengan tepat apakah sebuah ulasan mengandung sentimen positif atau negatif. Selain itu, metrik evaluasi seperti Classification Accuracy (82%), F1-Score (82%), dan Precision (86%) mengkonfirmasi bahwa model ini dapat dengan efisien membedakan antara sentimen positif dan negatif. Recall sebesar 82% juga menegaskan efektivitas model dalam menangkap sebagian besar ulasan positif. Hasil ini menunjukkan bahwa algoritma Naïve Bayes adalah alat yang sangat efektif untuk analisis sentimen, dengan keseimbangan yang baik dalam mendeteksi sentimen positif dan negatif. Meskipun kinerjanya sudah sangat baik, masih ada ruang untuk perbaikan, terutama dalam menangani ulasan netral atau ambigu. Secara keseluruhan, model Naïve Bayes terbukti menjadi metode yang efektif dan efisien untuk analisis sentimen pada konten yang dihasilkan pengguna, seperti ulasan di TikTok dan Tokopedia. Meskipun model Naïve Bayes telah menunjukkan tingkat akurasi yang tinggi, perbaikan lebih lanjut dapat dilakukan dengan mengeksplorasi algoritma machine learning yang lebih canggih, seperti Support Vector Machines (SVM) atau model deep learning, untuk melihat apakah mereka dapat mengungguli Naïve Bayes dalam hal precision dan recall. Selain itu, menangani masalah terkait ulasan yang ambigu atau netral dalam penelitian mendatang dapat meningkatkan kinerja model secara keseluruhan. Disarankan juga untuk memasukkan fitur tambahan, seperti kamus sentimen atau analisis berbasis konteks, guna memperbaiki klasifikasi sentimen dalam ulasan.

Terakhir, menerapkan model ini pada berbagai platform atau jenis konten yang dihasilkan pengguna lainnya dapat memberikan wawasan lebih lanjut mengenai generalisasi dan ketangguhan model dalam tugas analisis sentimen.

REFERENCES

[1] A. Septiani, A. Voutama, S. Siska, A. A. Hendriadi, and N. Heryana, "Analisis Sentimen Menggunakan Algoritma Naïve Bayes Terhadap Regulasi Tiktok Shop Pada Media Sosial X (Twitter)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 5729–5735, 2024.

[2] F. A. Indriyani, A. Fauzi, and S. Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023.

[3] M. Muljono, D. P. Artanti, A. Syukur, and A. Prihandono, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," *Konf. Nas. Sist. Inf. 2018*, 2018.

[4] R. P. Bastoni, "Analisis sentimen opini publik terhadap calon presiden indonesia 2024 menggunakan metode naive bayes dan lexicon based pada platform youtube," *Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta*.

[5] M. Taufik, "Analisis sentimen komentar Jasa Layanan Transportasi PT KCI dari Twitter Menggunakan Kombinasi Algoritma K-Nearest Neighbor (K-NN) dan Pendekatan Lexicon," 2018, *Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta*.

[6] G. Urva *et al.*, *PENERAPAN DATA MINING DI BERBAGAI BIDANG: Konsep, Metode, dan Studi Kasus*. PT. Sonpedia Publishing Indonesia, 2023.

[7] R. Safitri, N. Alfira, D. Tamitiadini, W. W. A. Dewi, and N. Febriani, *Analisis Sentimen: Metode Alternatif Penelitian Big Data*. Universitas Brawijaya Press, 2021.

[8] F. Arjunie, "Pengaruh Dukungan Produk Terhadap Perilaku Kelompok dan Sikap Merek: Dampak Langsung Terhadap Niat Pembelian Scarlett Whitening," 2024, *Universitas Islam Indonesia*.

[9] T. S. Wibowo *et al.*, "Transformasi Teknologi Komunikasi," 2024.

[10] I. C. N. Alamsyah and A. Apriyanti, "PENGARUH HEDONIC SHOPPING VALUE, WORD OF MOUTH DAN LIFESTYLE TERHADAP KEPUTUSAN PEMBELIAN SEPATU BEKAS PADA MAHASISWA DI KOTA MADIUN," in *SIMBA: Seminar Inovasi Manajemen, Bisnis, dan Akuntansi*, 2024.

[11] Armaeni, P. P., Wiguna, I. K. A. G., & Parwita, W. G. S. (2024). *Analisis Sentimen Komentar YouTube tentang Penutupan Toko TikTok Menggunakan Perbandingan Metode Naïve Bayes dan Decision Tree*. Jurnal Galaksi (Pengetahuan Global, Kecerdasan Buatan dan Sistem Informasi), 1(2), 70-80.

[12] Firdaus, A. A., Yudhana, A., & Riadi, I. (2023). *Analisis Opini Publik Calon Presiden Menggunakan Metode Naïve Bayes*. Kinetik: Teknologi Permainan, Sistem Informasi, Jaringan Komputer, Komputasi, Elektronika, dan Kontrol, 8(2), 563-570.

[13] Fadli, M., Triayudi, A., & Handayani, E. T. E. (2024). *Analisis Sentimen pada Aplikasi Media Sosial Twitter di Kenaikan Harga Bahan Bakar Minyak Menggunakan Naïve Bayes dan Decision Tree Algoritma*. SaNa: Jurnal Teknologi Blockchain, NFT, dan Metaverse, 2(2), 114-122.

[14] Kristanto, F., & Nugroho, A. Y. (2024). *Perbandingan Algoritma Naïve Bayes dan Pohon Keputusan Tentang Analisis Sentimen Data Komentar Siswa Di Era Digital Aplikasi Penilaian Guru (DITA)*. ULIL ALBAB: Jurnal Ilmiah Multidisiplin, 3(9), 1-12.

[15] Wahyuni, P., & Romli, M. A. (2024). *Perbandingan Algoritma Naïve Bayes Classifier dan Decision Tree untuk Analisis Sentimen terhadap Hak Angket DPR RI Twitter*. Jurnal Informatika dan Komputasi Terapan (JAIC), 8(2), 523-530.

[16] Wilim, N. N., & Oetama, R. S. (2021). *Analisis Sentimen tentang Pengacara Indonesia Program Televisi Klub Menggunakan K-Nearest Tetangga, Pengklasifikasi Naïve Bayes, dan Pohon Keputusan*. Jurnal Ilmiah Terapan, 8(1), 50-56.

[17] Zulkarnain, Mutia, R., Ariani, J. A., Barik, Z. A., & Azmi, H. (2024). *Perbandingan Performa K-Nearest Neighbor, Naive Bayes, dan Algoritma Pohon Keputusan untuk Klasifikasi Peringkat Netflix*. IJATIS: Jurnal Sains Terapan Teknologi dan Inovasi Indonesia, 1(1), 16-22.