



Evaluasi Komparatif Lima Model Machine Learning Untuk Klasifikasi Sentimen Ulasan Tokopedia Berbasis TF-IDF

Hana Kurnia Putri¹, Candhy Fadhila Arsyad², Muhammad Rifqi Fadhlan Rahmatullah³

Informatika, Fakultas Teknik, Universitas Doktor Nugroho Magetan

¹hanaputri2818@gmail.com, ²candhy.fadhila.arsyad@gmail.com, ³mrifqi.fadhlanr.90@gmail.com

Abstract

Marketplace application reviews contain information about transaction experiences, service quality, and technical issues, yet their large volume limits manual evaluation. This study comparatively evaluates five machine learning models for sentiment classification of Tokopedia reviews. Its contribution lies in testing the models under an identical TF-IDF representation scheme and interpreting the discriminative words and phrases produced by the best-performing model. A total of 5,000 reviews were collected from Google Play Store using the package identifier com.tokopedia.tkpdp. After removing neutral reviews and records that did not meet the processing criteria, the final dataset contained 4,695 reviews, comprising 2,195 negative and 2,500 positive reviews. The texts underwent cleaning, case folding, tokenization, stopword removal, and stemming before being divided using a stratified 80:20 train-test split. The evaluated models were Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, and Random Forest. Multinomial Naive Bayes achieved the best performance, with 89.78% accuracy, a macro F1-score of 89.76%, and a weighted F1-score of 89.79%. It correctly classified 404 negative and 439 positive reviews. Feature interpretation indicated that positive sentiment was strongly associated with terms such as mantap, keren, and bagus, whereas negative sentiment was characterized by terms such as sampah, jelek, and scam. These findings demonstrate that combining TF-IDF with Multinomial Naive Bayes provides an effective approach for mapping Tokopedia user perceptions.

Keywords: sentiment analysis, Tokopedia, TF-IDF, Multinomial Naive Bayes, machine learning

Abstrak

Ulasan pengguna aplikasi marketplace memuat informasi mengenai pengalaman transaksi, kualitas layanan, dan kendala teknis, tetapi volumenya menyulitkan evaluasi secara manual. Penelitian ini bertujuan mengevaluasi secara komparatif lima model *machine learning* untuk klasifikasi sentimen ulasan Tokopedia. Kontribusi penelitian terletak pada pengujian kelima model dalam satu skema representasi TF-IDF yang sama serta interpretasi kata dan frasa pembeda dari model terbaik. Sebanyak 5.000 ulasan dikumpulkan dari Google Play Store melalui identitas paket com.tokopedia.tkpdp. Setelah ulasan berlabel netral dan data yang tidak memenuhi kriteria dikeluarkan, diperoleh 4.695 ulasan, terdiri atas 2.195 ulasan negatif dan 2.500 ulasan positif. Data diproses melalui *cleaning*, *case folding*, tokenisasi, penghapusan *stopword*, dan *stemming*, kemudian dibagi secara *stratified* menjadi 80% data latih dan 20% data uji. Model yang dievaluasi meliputi Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest. Multinomial Naive Bayes menghasilkan performa terbaik dengan akurasi 89,78%, *macro F1* 89,76%, dan *weighted F1* 89,79%. Model tersebut mengklasifikasikan secara benar 404 ulasan negatif dan 439 ulasan positif. Interpretasi fitur menunjukkan bahwa sentimen positif didominasi kata “mantap”, “keren”, dan “bagus”, sedangkan sentimen negatif ditandai kata “sampah”, “jelek”, dan “scam”. Hasil ini menunjukkan bahwa kombinasi TF-IDF dan Multinomial Naive Bayes efektif untuk memetakan persepsi pengguna Tokopedia.

Kata kunci: analisis sentimen, Tokopedia, TF-IDF, Multinomial Naive Bayes, machine learning.



1. Pendahuluan

Perkembangan layanan digital telah memperluas fungsi aplikasi e-commerce dari sekadar sarana transaksi menjadi ruang interaksi antara pengguna dan penyedia platform. Melalui halaman ulasan Google Play Store, pengguna dapat menyampaikan pengalaman mengenai kemudahan penggunaan aplikasi, proses pembayaran, harga, pengiriman, layanan pelanggan, hingga tingkat kepercayaan terhadap platform. Informasi tersebut dapat menjadi sumber evaluasi karena berasal dari pengalaman langsung pengguna. Namun, jumlah ulasan yang terus bertambah membuat pembacaan dan pengelompokan opini secara manual menjadi tidak efisien. Penelitian terdahulu menunjukkan bahwa ulasan Google Play Store telah banyak dimanfaatkan untuk menganalisis persepsi pengguna pada aplikasi komunitas digital, layanan publik, game, media sosial, perbankan digital, dan e-commerce [1], [2], [3], [4].

Analisis sentimen merupakan bagian dari pemrosesan bahasa alami yang digunakan untuk menentukan kecenderungan opini dalam teks, misalnya positif atau negatif. Pada data ulasan aplikasi, pendekatan ini memungkinkan komentar tidak terstruktur diubah menjadi informasi yang dapat dihitung dan dibandingkan. Prosesnya umumnya mencakup pengumpulan ulasan, penetapan label, pembersihan teks, pembentukan fitur numerik, pelatihan model, dan evaluasi klasifikasi. Berbagai penelitian pada aplikasi Jamsostek Mobile, CapCut, Money Lover, Kitabisa, TikTok, AdaKami, KitaLulus, GoBiz, INFO BMKG, Roblox, MyBCA, SEA Bank, dan aplikasi operator seluler memperlihatkan bahwa model *machine learning* klasik masih relevan untuk mengolah ulasan berbahasa Indonesia karena kebutuhan komputasinya relatif ringan dan performanya tetap kompetitif [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17].

Teks ulasan perlu direpresentasikan dalam bentuk numerik sebelum diproses oleh model klasifikasi. Penelitian ini menggunakan *Term Frequency-Inverse Document Frequency* atau TF-IDF, yaitu pembobotan yang mempertimbangkan frekuensi suatu istilah dalam dokumen sekaligus tingkat kelangkaannya dalam keseluruhan korpus. Kata yang sering muncul pada satu ulasan, tetapi tidak terlalu umum pada seluruh data, akan memperoleh bobot yang lebih informatif. Representasi ini menghasilkan matriks fitur berdimensi tinggi dan bersifat *sparse*, sehingga sesuai digunakan sebagai masukan bagi berbagai model klasifikasi teks. TF-IDF juga masih banyak digunakan sebagai *baseline* dalam penelitian ulasan Tokopedia, Shopee, Bukalapak, dan ulasan pelanggan

dari domain lain [18], [19], [20], [21], [22], [23], [24], [25], [26].

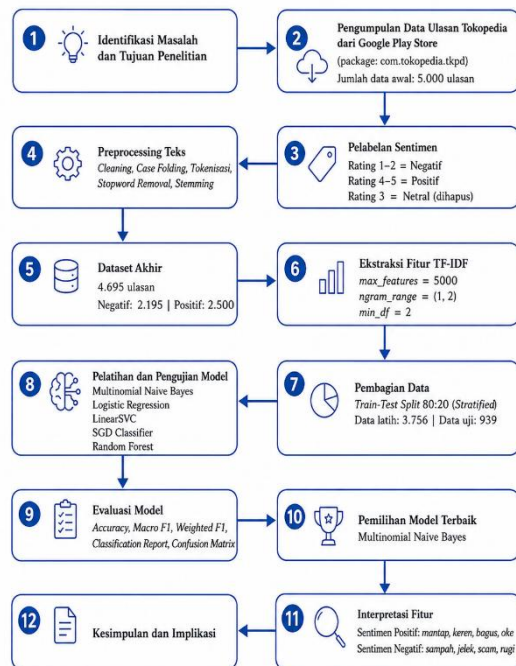
Meskipun analisis sentimen terhadap ulasan aplikasi telah banyak dilakukan, masih ditemukan beberapa keterbatasan pada penelitian terdahulu. Sebagian kajian berfokus pada aplikasi non-marketplace, seperti layanan publik, media sosial, game, hiburan, dan perbankan digital. Sejumlah penelitian juga hanya membandingkan dua atau tiga algoritma, sehingga belum memberikan gambaran yang cukup luas mengenai perilaku model probabilistik, linear, optimisasi stokastik, dan *ensemble* ketika menerima representasi fitur yang identik [1], [2], [3], [4], [27], [28], [29], [30]. Selain itu, banyak penelitian berhenti pada penyajian metrik evaluasi tanpa menguraikan kata atau frasa yang menjadi pembeda sentimen pada model terbaik. Kondisi tersebut menunjukkan perlunya pengujian yang tidak hanya membandingkan skor klasifikasi, tetapi juga menjelaskan pola linguistik yang membentuk hasil prediksi.

Berdasarkan kesenjangan tersebut, penelitian ini mengklasifikasikan sentimen ulasan aplikasi Tokopedia yang dikumpulkan dari Google Play Store melalui paket `com.tokopedia.tkp`. TF-IDF digunakan untuk membentuk fitur unigram dan bigram, kemudian lima model, yaitu Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest, dievaluasi pada dataset dan skema validasi yang sama. Kontribusi utama penelitian ini adalah menyediakan evaluasi komparatif terhadap lima model yang mewakili karakteristik algoritmik berbeda dalam satu lingkungan eksperimen yang konsisten, sekaligus menginterpretasikan kata dan frasa yang paling membedakan sentimen positif dan negatif pada model terbaik. Dengan pendekatan tersebut, penelitian ini tidak hanya menentukan model yang paling sesuai untuk data ulasan Tokopedia, tetapi juga menghasilkan informasi yang dapat digunakan untuk memahami pola persepsi pengguna terhadap layanan e-commerce.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif berbasis eksperimen untuk menguji performa beberapa model machine learning dalam klasifikasi sentimen ulasan aplikasi Tokopedia. Eksperimen dilakukan sesuai gambar 1, yang dimulai dari pengambilan data, pemberian label sentimen, pengolahan teks, pembentukan fitur numerik, pelatihan model, hingga evaluasi hasil klasifikasi. Pendekatan ini dipilih karena memungkinkan setiap model diuji pada data dan skema evaluasi yang sama,

sehingga perbandingan performa dapat dilakukan secara lebih objektif [5], [18], [20], [28].



Gambar 1. Diagram alur penelitian analisis sentimen ulasan Tokopedia

2.1. Pengumpulan Data

Data penelitian berupa ulasan pengguna aplikasi Tokopedia yang tersedia pada Google Play Store. Pengumpulan ulasan dilakukan melalui identitas paket aplikasi `com.tokopedia.tkpdp` untuk memastikan bahwa data berasal dari halaman resmi Tokopedia. Informasi yang diambil meliputi teks ulasan dan skor rating yang diberikan pengguna, dengan jumlah awal sebanyak 5.000 ulasan. Seluruh data kemudian disimpan dalam format CSV agar dapat diproses lebih lanjut menggunakan Python. Ulasan Google Play Store dipilih karena memuat pengalaman langsung pengguna ketika menggunakan fitur dan layanan aplikasi [2], [3], [4], [9], [27].

2.2. Pelabelan Sentimen

Pelabelan sentimen dilakukan menggunakan skor rating yang menyertai setiap ulasan. Ulasan dengan rating 1 dan 2 dikategorikan sebagai sentimen negatif, sedangkan rating 4 dan 5 dikategorikan sebagai sentimen positif. Ulasan dengan rating 3 dianggap netral dan dikeluarkan dari dataset karena penelitian ini menggunakan klasifikasi biner serta untuk mengurangi kemungkinan ambiguitas antara sentimen positif dan negatif. Setelah proses seleksi tersebut dan penghapusan data yang tidak memenuhi kriteria, diperoleh dataset akhir sebanyak 4.695 ulasan, yang terdiri atas 2.195 ulasan negatif dan 2.500 ulasan positif. Skema pelabelan berbasis rating digunakan karena skor yang diberikan pengguna dapat menjadi indikator awal terhadap polaritas

pengalaman yang disampaikan dalam ulasan [2], [3], [4], [9], [27].

2.3. Preprocessing Text

Tahap preprocessing dilakukan untuk mengubah teks ulasan mentah menjadi data yang lebih bersih dan siap diproses. Proses ini meliputi penghapusan karakter yang tidak diperlukan, penyamaan huruf menjadi huruf kecil, pemecahan kalimat menjadi token, penghapusan stopwords, dan stemming bahasa Indonesia. Pada tahap penghapusan stopwords, kata negasi seperti “tidak”, “bukan”, “belum”, “kurang”, dan bentuk tidak baku sejenisnya tetap dipertahankan karena dapat mengubah makna sentimen dalam kalimat. Misalnya, frasa “bagus” dan “tidak bagus” memiliki polaritas yang berbeda, sehingga kata negasi tidak boleh dihapus. Tahap ini dilakukan untuk mengurangi noise pada teks sekaligus mempertahankan kata-kata yang memiliki pengaruh terhadap polaritas ulasan [14], [16], [17], [29].

2.4. Ekstraksi Fitur TF-IDF

Setelah teks diproses, data ulasan dikonversi ke dalam format numerik dengan metode Term Frequency-Inverse Document Frequency (TF-IDF). Dalam penelitian ini, TF-IDF diterapkan dengan parameter `max_features=5000`, `ngram_range=(1,2)`, dan `min_df=2`. Parameter `max_features=5000` digunakan untuk membatasi jumlah fitur guna memastikan efisiensi komputasi. Penggunaan `ngram_range=(1,2)` memungkinkan model menangkap kata tunggal dan frasa dua kata, sedangkan `min_df=2` digunakan untuk mengabaikan kata-kata yang terlalu jarang muncul. Kombinasi parameter ini dipilih agar representasi teks dapat menangkap pola-pola penting dalam ulasan, seperti “aplikasi bagus,” “tidak bisa,” “login gagal,” atau “mahal.” TF-IDF digunakan karena metode ini masih menjadi metode representasi teks yang banyak digunakan dalam penelitian analisis sentimen berdasarkan model klasik [18], [20], [25], [26], [29].

2.5. Model Klasifikasi dan Evaluasi

Penelitian ini membandingkan lima model klasifikasi: Multinomial Naive Bayes, Regresi Logistik, LinearSVC, SGD Classifier, dan Random Forest. Kelima model ini dipilih karena sering digunakan dalam klasifikasi teks dan dapat berfungsi sebagai tolok ukur untuk mengidentifikasi model yang paling sesuai untuk data ulasan Tokopedia. Data, yang telah direpresentasikan menggunakan TF-IDF, kemudian dibagi menjadi set pelatihan dan set pengujian. Setiap model dilatih menggunakan data pelatihan dan kemudian diuji menggunakan data pengujian yang sama untuk memastikan perbandingan hasil evaluasi yang adil.

Evaluasi kinerja dilakukan menggunakan accuracy, macro F1, weighted F1, classification report, dan confusion matrix. Akurasi digunakan untuk menentukan persentase keseluruhan prediksi yang benar. Macro F1 digunakan untuk menilai keseimbangan kinerja di setiap kelas tanpa

mempertimbangkan jumlah data di setiap kelas, sedangkan weighted F1 memperhitungkan proporsi data di setiap kelas. Sementara itu, matriks kebingungan digunakan untuk memeriksa detail prediksi yang benar dan salah baik di kelas positif maupun negatif. Model dengan kinerja terbaik kemudian dianalisis lebih lanjut melalui interpretasi fitur untuk mengidentifikasi kata atau frasa yang paling erat kaitannya dengan sentimen positif dan negatif. [1], [13], [28], [30].

3. Hasil dan Pembahasan

Pengujian dilakukan terhadap lima model machine learning dengan representasi fitur TF-IDF. Model yang diuji meliputi Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest. Berdasarkan hasil evaluasi, seluruh model menunjukkan performa yang cukup baik karena memperoleh nilai accuracy di atas 87%. Model dengan performa terbaik adalah Multinomial Naive Bayes dengan accuracy sebesar 0,897764, macro F1 sebesar 0,897621, dan weighted F1 sebesar 0,897869. Hasil evaluasi model terbaik ditampilkan pada gambar 2.

	precision	recall	f1-score	support
Negatif	0.87	0.92	0.89	439
Positif	0.93	0.88	0.90	500
accuracy			0.90	939
macro avg	0.90	0.90	0.90	939
weighted avg	0.90	0.90	0.90	939

Gambar 2. Hasil model Multinomial Naive Bayes

SVM/LinearSVC menempati peringkat kedua dengan akurasi sebesar 0,896699, skor F1 makro sebesar 0,896405, dan weighted F1 sebesar 0,896763. Regresi Logistik menempati peringkat ketiga dengan akurasi sebesar 0,894569, skor F1 makro sebesar 0,894446, dan skor F1 tertimbang sebesar 0,894680. Sementara itu, SGD Classifier dan Random Forest mencatat skor kinerja yang lebih rendah dibandingkan dengan ketiga model lainnya. Ringkasan perbandingan kinerja semua model ditampilkan pada Tabel 1.

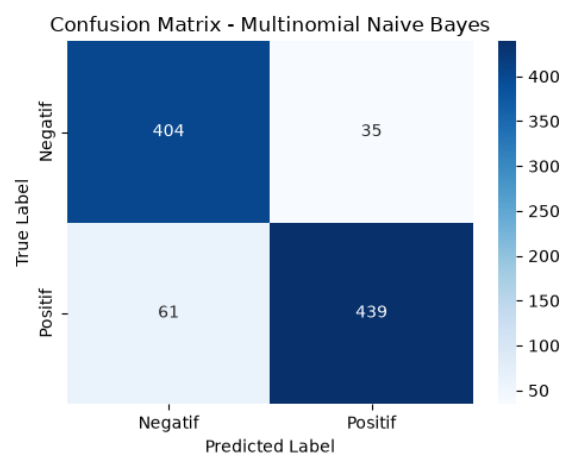
Tabel 1. Perbandingan performa model

Model	Accuracy	Macro F1	Weighted F1
Multinomial Naive Bayes	0,897764	0,897621	0,897869
SVM / LinearSVC	0,896699	0,896405	0,896763
Logistic Regression	0,894569	0,894446	0,894680
SGD Classifier	0,884984	0,884719	0,885078
Random Forest	0,876464	0,876346	0,876594

Hasil tersebut menunjukkan bahwa Multinomial Naive Bayes memiliki performa paling stabil dibandingkan model lainnya. Perbedaan performa antara Multinomial Naive Bayes, SVM/LinearSVC,

dan Logistic Regression memang tidak terlalu besar, tetapi Multinomial Naive Bayes tetap unggul pada tiga metrik utama yang digunakan, yaitu accuracy, macro F1, dan weighted F1. Oleh karena itu, model tersebut dipilih sebagai model terbaik untuk dianalisis lebih lanjut.

Berdasarkan confusion matrix untuk model Multinomial Naive Bayes, 404 ulasan negatif diklasifikasikan dengan benar sebagai negatif. Selain itu, 439 ulasan positif juga diklasifikasikan dengan benar sebagai positif. Kesalahan klasifikasi terjadi pada 35 ulasan negatif yang diprediksi sebagai positif dan 61 ulasan positif yang diprediksi sebagai negatif. Dengan demikian, terdapat 843 prediksi yang benar dari total 939 titik data uji. Visualisasi confusion matrix untuk model terbaik ditunjukkan pada Gambar 3.



Gambar 3. Confusion matrix model Multinomial Naive Bayes

Perbandingan confusion matrix pada seluruh model menunjukkan bahwa setiap algoritma memiliki pola kesalahan klasifikasi yang berbeda. SVM/LinearSVC mampu mengenali ulasan positif paling banyak, yaitu 446 data positif yang diprediksi dengan benar. Namun, model tersebut juga menghasilkan 43 ulasan negatif yang salah diprediksi sebagai positif. Multinomial Naive Bayes memiliki jumlah kesalahan yang lebih seimbang, yaitu 35 false positive dan 61 false negative. Ringkasan nilai confusion matrix setiap model ditampilkan pada Tabel 2.

Tabel 2. Ringkasan confusion matrix seluruh model

Model	TN	FP	FN	TP
Multinomial Naive Bayes	404	35	61	439
Logistic Regression	404	35	64	436
SVM / LinearSVC	396	43	54	446
SGD Classifier	393	46	62	438
Random Forest	397	42	74	426

Hasil pengujian menunjukkan bahwa Multinomial Naive Bayes menjadi model dengan performa paling tinggi dalam klasifikasi sentimen ulasan pengguna Tokopedia. Model ini memperoleh nilai accuracy, macro F1, dan weighted F1 yang lebih baik

dibandingkan Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest. Keunggulan tersebut dapat terjadi karena karakteristik Multinomial Naive Bayes sesuai dengan data teks yang direpresentasikan menggunakan TF-IDF. Pada data ulasan yang berbentuk teks pendek, pola kemunculan kata atau frasa tertentu sering kali sudah cukup kuat untuk membedakan sentimen positif dan negatif. Oleh karena itu, model probabilistik seperti Naive Bayes masih relevan digunakan sebagai pendekatan dasar dalam analisis sentimen berbasis teks [5], [7], [19], [23], [24].

Jika dibandingkan dengan lima penelitian pembandingan utama, hasil penelitian ini menunjukkan peningkatan performa yang cukup jelas. Penelitian Weverse memperoleh akurasi terbaik 83% dengan Random Forest, sedangkan Naive Bayes memperoleh 74,2% [1]. Penelitian PLN Mobile menunjukkan bahwa Multinomial Naive Bayes memperoleh akurasi 70,89% [2]. Penelitian PUBG Mobile memperoleh performa sekitar 70% pada model Naive Bayes dan SVM [3]. Penelitian aplikasi X dan BCA Mobile juga menunjukkan bahwa model Naive Bayes dan SVM masih sering digunakan untuk ulasan aplikasi, tetapi performanya sangat dipengaruhi oleh karakteristik data, pelabelan, dan distribusi kelas [4], [27].

Keunggulan penelitian ini terletak pada penggunaan lima model pembandingan dan interpretasi fitur pada model terbaik. Beberapa penelitian terdahulu umumnya hanya membandingkan dua atau tiga model, seperti Naive Bayes dan SVM, atau menambahkan satu model lain sebagai pembandingan [13], [16], [17], [28]. Dengan membandingkan Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest, penelitian ini memberikan gambaran yang lebih luas mengenai performa model machine learning klasik pada data ulasan Tokopedia.

4. Kesimpulan

Penelitian ini mengevaluasi lima model *machine learning* untuk klasifikasi sentimen ulasan Tokopedia menggunakan representasi TF-IDF, yaitu Multinomial Naive Bayes, Logistic Regression, LinearSVC, SGD Classifier, dan Random Forest. Hasil pengujian menunjukkan bahwa Multinomial Naive Bayes memberikan performa terbaik dengan *accuracy* sebesar 89,78%, *macro F1* sebesar 89,76%, dan *weighted F1* sebesar 89,79%. Model tersebut mampu mengklasifikasikan sentimen positif dan negatif secara relatif seimbang, meskipun masih ditemukan kesalahan pada ulasan yang mengandung bahasa tidak baku, konteks ambigu, atau ketidaksesuaian antara teks dan rating.

Kontribusi ilmiah penelitian ini mencakup tiga aspek. Dalam bidang *sentiment analysis*, penelitian ini menunjukkan bahwa kombinasi TF-IDF berbasis unigram dan bigram dengan interpretasi fitur dapat

digunakan untuk mengidentifikasi polaritas sekaligus menemukan kata dan frasa yang menjadi penanda utama sentimen pengguna. Dalam bidang *machine learning*, penelitian ini menyediakan evaluasi komparatif terhadap model probabilistik, model linear, optimisasi stokastik, dan model *ensemble* pada dataset, representasi fitur, serta skema pengujian yang sama. Hasil tersebut memperlihatkan bahwa Multinomial Naive Bayes lebih sesuai untuk karakteristik ulasan Tokopedia yang direpresentasikan dalam bentuk matriks TF-IDF berdimensi tinggi dan bersifat *sparse*.

Dalam konteks *e-commerce analytics*, hasil klasifikasi dan interpretasi fitur mampu mengubah ulasan tidak terstruktur menjadi informasi yang lebih mudah dianalisis. Kata-kata positif seperti “mantap”, “keren”, “bagus”, dan “oke” menunjukkan apresiasi pengguna, sedangkan kata-kata negatif seperti “sampah”, “jelek”, “scam”, “batal”, dan “rugi” mengindikasikan masalah kepuasan, transaksi, kualitas layanan, dan kepercayaan terhadap platform. Dengan demikian, pendekatan yang dihasilkan dapat menjadi dasar bagi pengelola aplikasi e-commerce untuk memantau persepsi pengguna, mengidentifikasi isu layanan, dan menentukan prioritas perbaikan secara lebih terukur.

Ucapan Terimakasih

Ucapan terimakasih penulis sampaikan kepada seluruh pihak yang telah memberikan dukungan dalam proses penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada Universitas Doktor Nugroho Magetan atas dukungan akademik yang diberikan sehingga penelitian ini dapat diselesaikan dengan baik.

Daftar Rujukan

- [1] A. Fitriyani and I. Ibrahim, “Performance Comparison of Naïve Bayes, Random Forest, and Support Vector Machine Algorithms in Sentiment Analysis of the Weverse Application,” *Sist. J. Sist. Inf.*, 2026, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5893>
- [2] J. M. Ayomi, A. V. Vitianingsih, Y. Kristyawan, A. L. Maukar, and T. Widiartin, “Sentiment Analysis of User Reviews for the PLN Mobile Application Using Naïve Bayes and Long Short-Term Memory,” *J. Inf. Syst. Informatics*, vol. 7, no. 4, 2025, doi: 10.63158/journalisi.v7i4.1342.
- [3] P. R. Sari, D. R. Indah, E. Rasywir, M. A. Firdaus, and G. Athalina, “Comparison of Naive Bayes and SVM Algorithms for Sentiment Analysis of PUBG Mobile on Google Play Store,” *Sist. J. Sist. Inf.*, vol. 13, no. 6, pp. 2767–2779, 2024, doi: 10.32520/stmsi.v13i6.4814.
- [4] Eskiaturrofikoh and R. R. Suryono, “Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine (SVM),” *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1408–1419, 2024, doi: 10.29100/jipi.v9i3.5392.

- [5] T. Setyani, K. Sari, H. N. Heidy, and R. R. Suryono, "Analisis Sentimen Pengguna Aplikasi Jamsostek Mobile Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support Vector Machine dan Naive Bayes," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. 1, 2026, doi: 10.57152/malcom.v6i1.2526.
- [6] P. Setiaji, "Sentiment Analysis of CapCut Application Reviews using Support Vector Machine and SMOTE," *Sist. J. Sist. Inf.*, 2026, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5948>
- [7] N. A. S. Bila, "Sentiment Analysis of Money Lover App Reviews using Random Forest and Naive Bayes," *Sist. J. Sist. Inf.*, 2026, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5859>
- [8] N. P. Maharani, "Analisis Sentimen Ulasan Pengguna Pada Aplikasi Kitabisa Donasi Zakat Menggunakan Metode Support Vector Machine (SVM) dan Naive Bayes," *J. Algoritma.*, 2026, [Online]. Available: <https://jurnal.itg.ac.id/index.php/algoritma/article/view/3376>
- [9] M. T. F. Maulana, "Analisis Sentimen Ulasan Aplikasi TikTok di Google Playstore Menggunakan Algoritma Naive Bayes," *RIGGS J. Artif. Intell. Digit. Bus.*, 2025, [Online]. Available: <https://journal.ilmudata.co.id/index.php/RIGGS/article/view/2962>
- [10] A. S. Mubarakah, "Penerapan Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Pengguna Aplikasi AdaKami di Google Play Store," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.6191.
- [11] A. S. A. Rafsanjani, D. L. Fithri, and S. Supriyono, "Sentiment Analysis of User Reviews of the KitaLulus Application on Google Play Store using the Support Vector Machine (SVM) Algorithm," *Sist. J. Sist. Inf.*, 2025, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5519>
- [12] H. Firda, P. Putra, N. R. Oktadini, and P. E. Sevtiyuni, "Comparison of Rating-based and Inset Lexicon-based Labeling in Sentiment Analysis using SVM: A Case Study of GoBiz Reviews on Google Play Store," *Sist. J. Sist. Inf.*, vol. 14, no. 2, pp. 516–528, 2025, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/4795>
- [13] N. Aditiya, P. Setiaji, and Supriyono, "Analisis Sentimen Kepuasan Masyarakat terhadap Aplikasi INFO BMKG menggunakan Naive Bayes, SVM, dan KNN," *Sist. J. Sist. Inf.*, vol. 14, no. 3, pp. 1418–1432, 2025, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5223>
- [14] N. K. F. P. Dewi, "Sentiment Analysis of Roblox Game Reviews Using SVM with TF-IDF Weighting and Lexicon InSet," *Sink. J. dan Penelit. Tek. Inform.*, 2025, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/15272>
- [15] M. R. M. Akbar, "Sentiment Analysis of MyBCA Application User Reviews using Naive Bayes, Random Forest, and Decision Tree," *Sist. J. Sist. Inf.*, 2025, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/5472>
- [16] Y. Christian, "Sentiment Analysis by Using Naive Bayes Classification and Support Vector Machine on SEA Bank Application Reviews," *Sink. J. dan Penelit. Tek. Inform.*, 2024, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/13141>
- [17] T. S. Ningsih, "Sentiment Analysis of Mobile Provider Application Reviews Using Naive Bayes Algorithm and Support Vector Machine," *Sink. J. dan Penelit. Tek. Inform.*, 2024, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/13469>
- [18] D. Nurmadewi, Z. F. Jailani, H. Rafi, D. A. Anggoro, and D. Setiowati, "Perbandingan Support Vector Machine dan Naive Bayes untuk Klasifikasi Sentimen Ulasan E-Commerce," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 6, no. 2, pp. 923–933, 2026, doi: 10.57152/malcom.v6i2.2648.
- [19] R. Tajuddin, "Classification of Sentiment Tokopedia and Shopee App Reviews Using Naive Bayes Algorithm," *Bit-Tech.*, 2025, [Online]. Available: <https://jurnal.kdi.or.id/index.php/bt/article/view/3246>
- [20] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," *J. Appl. Informatics Comput.*, vol. 9, no. 5, pp. 2552–2561, 2025, doi: 10.30871/jaic.v9i5.10524.
- [21] M. Idris, "Sentiment Analysis of Tokopedia App Reviews using Word2Vec and FastText Feature Extraction with Machine Learning Algorithms," *Sink. J. dan Penelit. Tek. Inform.*, 2025, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/14278>
- [22] S. F. Kadir, "Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Machine Learning," *RIGGS J. Artif. Intell. Digit. Bus.*, 2025, [Online]. Available: <https://journal.ilmudata.co.id/index.php/RIGGS/article/view/2850>
- [23] R. R. Rismansyah, "Analisis Sentimen Ulasan Shopee pada Google Play Store Menggunakan Algoritma Naive Bayes," *JEIS J. Elektro dan Inform. Swadharma*, 2025, doi: 10.56486/jeis.vol5no1.661.
- [24] M. D. C. Fardani, "Application of Naive Bayes for Sentiment Analysis of Shopee User Reviews," *Bit-Tech.*, 2025, [Online]. Available: <https://jurnal.kdi.or.id/index.php/bt/article/view/2854>
- [25] F. Akmal, "Optimization Naive Bayes Algorithm in Sentiment Analysis of Bukalapak Application Reviews," *Sink. J. dan Penelit. Tek. Inform.*, 2024, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/sinkron/article/view/13132>
- [26] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, p. 340, 2024, doi: 10.3390/computers13120340.
- [27] D. W. Bhatara and R. R. Suryono, "Analisis Sentimen Aplikasi BCA Mobile Menggunakan Algoritma Naive Bayes Dan Support Vector Machine," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 4, pp. 1907–1917, 2024, doi: 10.29100/jupi.v9i4.5536.
- [28] S. Jatmiko and C. Dometian, "Analisis Sentimen Ulasan Google Play Store: Studi Komparatif Algoritma SVM, Naive Bayes, dan Logistic Regression," *J. Fasilkom*, vol. 15, no. 3, 2025, doi: 10.37859/jf.v15i3.10016.
- [29] S. Khoerunnisa, D. F. Shiddieq, and D. Nurhayati, "Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 2, pp. 566–577, 2025, doi: 10.57152/malcom.v5i2.1852.

- [30] R. Alawaji and A. Aloraini, "Sentiment Analysis of Digital Banking Reviews Using Machine Learning and Large Language Models," *Electronics*, vol. 14, no. 11, p. 2125, 2025, doi: 10.3390/electronics14112125.