

PENINGKATAN MODEL KLASIFIKASI SENTIMEN ULASAN APLIKASI ALFAGIFT MENGGUNAKAN NAIVE BAYES.

Fauzy¹, Nining Rahaningsih², Irfan Ali³, Mulyawan⁴.

Program Studi Teknik Informatika¹
Program Studi Komputerisasi Akuntansi²
Program Studi Rekayasa Perangkat Lunak³
Program Studi Sistem Informasi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
afauzy0908@gmail.com

(*) Corresponding Author : afauzy0908@gmail.com
Published : 15 July 2026

Abstract—This study aims to improve the performance of sentiment classification models on Alfagift application reviews using the Naive Bayes algorithm. The dataset was collected through scraping from the Google Play Store, resulting in 10,000 reviews, which were then cleaned to obtain 8,510 valid data entries. The research process includes text preprocessing consisting of cleaning, case folding, word normalization, tokenizing, stopword removal, and stemming, as well as feature extraction using the TF-IDF method. Sentiment labeling was conducted using a lexicon-based approach, classifying the data into three categories: positive, negative, and neutral. The classification model was developed in two scenarios: a baseline model using CountVectorizer and an improved model using TF-IDF with n-gram (1,2) along with alpha parameter tuning in the Multinomial Naive Bayes algorithm. Model evaluation was performed using accuracy, precision, recall, and F1-score metrics. The results show that the baseline model achieved an accuracy of 0.679, precision of 0.657, recall of 0.595, and F1-score of 0.608. After applying improvements, the model performance increased, achieving an accuracy of 0.722, precision of 0.705, recall of 0.653, and F1-score of 0.669. Furthermore, the sentiment distribution analysis indicates that positive sentiment dominates the dataset, followed by neutral and negative sentiments. The most frequently occurring words in the reviews relate to promotions, shopping, application features, and delivery, reflecting users' experiences with the Alfagift service. Based on these findings, it can be concluded that comprehensive preprocessing, the use of TF-IDF with n-gram, and parameter tuning in Naive Bayes significantly improve sentiment classification performance.

Keywords: Sentiment Analysis, Alfagift, Naive Bayes, TF-IDF, Preprocessing, Model Optimization.

Abstrak—Penelitian ini bertujuan untuk meningkatkan performa model klasifikasi sentimen pada ulasan aplikasi Alfagift dengan menggunakan algoritma Naive Bayes. Data yang digunakan dalam penelitian ini diperoleh melalui proses scraping dari Google Play Store sebanyak 10.000 ulasan, yang kemudian melalui tahap pembersihan data sehingga diperoleh 8.510 data yang valid. Tahapan penelitian meliputi preprocessing teks yang terdiri dari cleaning, case folding, normalisasi kata, tokenizing, stopword removal, dan stemming, serta ekstraksi fitur menggunakan metode TF-IDF. Pelabelan sentimen dilakukan menggunakan pendekatan lexicon-based dengan mengklasifikasikan data ke dalam tiga kategori, yaitu positif, negatif, dan netral. Model klasifikasi dibangun dalam dua skenario, yaitu model baseline menggunakan CountVectorizer dan model improved menggunakan TF-IDF dengan n-gram (1,2) serta tuning parameter alpha pada algoritma Multinomial Naive Bayes. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa model baseline menghasilkan accuracy sebesar 0,679, precision 0,657, recall 0,595, dan F1-score 0,608. Setelah dilakukan peningkatan, model improved menunjukkan peningkatan performa dengan accuracy sebesar 0,722, precision 0,705, recall 0,653, dan F1-score 0,669. Selain itu, hasil analisis menunjukkan bahwa distribusi sentimen didominasi oleh sentimen positif, diikuti oleh netral dan negatif. Kata-kata dominan yang muncul dalam ulasan berkaitan dengan promo, belanja, aplikasi, dan pengiriman, yang mencerminkan pengalaman pengguna terhadap layanan Alfagift. Berdasarkan hasil tersebut, dapat disimpulkan bahwa penerapan preprocessing yang komprehensif, penggunaan TF-IDF dengan n-gram, serta tuning parameter pada Naive Bayes mampu meningkatkan performa klasifikasi sentimen secara signifikan.

Kata Kunci: Analisis Sentimen, Alfagift, Naive Bayes, TF-IDF, Preprocessing, Optimasi Model.

INTRODUCTION

Perkembangan teknologi informasi dan komunikasi yang sangat pesat telah mengubah pola interaksi antara perusahaan dan konsumen di Indonesia. Saat ini, aplikasi mobile e-commerce menjadi salah satu platform utama yang digunakan jutaan masyarakat untuk berbelanja secara digital. Aplikasi Alfagift sebagai platform resmi milik PT Sumber Alfaria Trijaya Tbk (Alfamart) telah menjadi salah satu aplikasi ritel modern yang paling banyak diunduh dan digunakan oleh pelanggan [1]. Dengan semakin tingginya jumlah pengguna, aplikasi ini menghasilkan volume ulasan pengguna yang sangat besar di Google Play Store setiap harinya. Ulasan tersebut tidak hanya mencerminkan pengalaman pengguna, tetapi juga menjadi sumber data penting bagi perusahaan untuk mengevaluasi kualitas layanan, fitur aplikasi, dan tingkat kepuasan pelanggan secara keseluruhan.

Ulasan pengguna di Google Play Store mengandung opini, keluhan, pujian, serta saran yang sangat beragam dan tidak terstruktur. Data teks ulasan ini menjadi aset berharga bagi perusahaan e-commerce seperti Alfagift untuk memahami kebutuhan dan persepsi pelanggan secara real-time [2]. Namun, memproses ribuan hingga puluhan ribu ulasan secara manual sangatlah tidak efisien dan memakan waktu. Oleh karena itu, diperlukan pendekatan otomatis berbasis kecerdasan buatan untuk menganalisis sentimen ulasan tersebut. Analisis sentimen merupakan cabang penting dari Natural Language Processing (NLP) yang mampu mengklasifikasikan teks menjadi kategori positif, negatif, atau netral dengan cepat dan akurat.

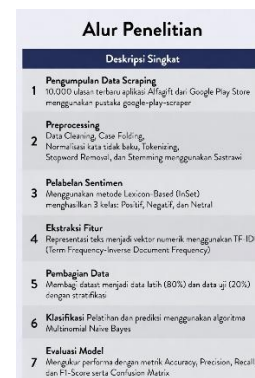
Salah satu algoritma klasifikasi yang paling sering digunakan dalam analisis sentimen adalah Naive Bayes. Algoritma ini dipilih karena memiliki keunggulan berupa kecepatan komputasi yang tinggi, kesederhanaan implementasi, serta kemampuan menangani data teks berdimensi tinggi [3]. Meskipun demikian, performa Naive Bayes pada data ulasan aplikasi e-commerce berbahasa Indonesia masih sering suboptimal. Hal ini disebabkan oleh karakteristik bahasa Indonesia yang kaya akan slang, singkatan, emoji, dan variasi penulisan yang tidak baku. Tanpa preprocessing yang komprehensif dan pemilihan fitur yang tepat, akurasi model Naive Bayes dasar pada ulasan aplikasi serupa biasanya hanya berada di kisaran 55–60%.

Penelitian terdahulu menunjukkan bahwa akurasi Naive Bayes dapat ditingkatkan secara signifikan melalui kombinasi teknik preprocessing (cleaning, case folding, tokenizing, stopword removal, dan stemming) serta metode ekstraksi fitur TF-IDF. Selain itu, tuning parameter alpha pada Multinomial Naive Bayes juga terbukti mampu meningkatkan performa model [4]. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan fokus pada peningkatan model klasifikasi sentimen khusus untuk ulasan aplikasi Alfagift. Dengan mengoptimalkan seluruh proses tersebut, diharapkan akurasi klasifikasi dapat meningkat dan memberikan hasil yang lebih akurat dibandingkan model baseline [5].

Hasil dari penelitian ini tidak hanya bermanfaat secara akademis dalam pengembangan ilmu Natural Language Processing dan Machine Learning di Indonesia, tetapi juga secara praktis bagi pengelola aplikasi Alfagift. Model yang dihasilkan dapat menjadi alat analitik yang membantu perusahaan dalam memantau sentimen pelanggan secara cepat, mengidentifikasi masalah teknis, serta merencanakan perbaikan fitur dan layanan [6]. Dengan demikian, penelitian ini diharapkan dapat berkontribusi pada peningkatan kualitas aplikasi Alfagift serta kepuasan pengguna secara keseluruhan.

MATERIALS AND METHODS

Penelitian ini mengikuti alur metodologi yang sistematis dan linier untuk memastikan setiap tahapan saling mendukung dalam mencapai tujuan peningkatan model klasifikasi sentimen. Seluruh proses diimplementasikan menggunakan bahasa pemrograman Python pada platform Google Colab.



Gambar 1 Alur Penelitian

Gambar tersebut menggambarkan alur penelitian analisis sentimen yang dilakukan secara

sistematis dalam tujuh tahapan utama. Penelitian diawali dengan pengumpulan data melalui teknik scraping sebanyak 10.000 ulasan aplikasi Alfagift dari Google Play Store menggunakan pustaka *google-play-scraper*. Selanjutnya dilakukan tahap *preprocessing* yang meliputi pembersihan data, *case folding*, normalisasi kata tidak baku, tokenizing, penghapusan *stopword*, serta *stemming* menggunakan Sastrawi. Setelah itu, data diberi label sentimen menggunakan metode *lexicon-based* (InSet) yang mengelompokkan data ke dalam tiga kelas, yaitu positif, negatif, dan netral.

Tahapan berikutnya adalah ekstraksi fitur dengan mengubah teks menjadi representasi numerik menggunakan metode TF-IDF. Data kemudian dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% dengan teknik stratifikasi. Pada tahap klasifikasi, model dilatih dan digunakan untuk prediksi menggunakan algoritma Multinomial Naive Bayes. Terakhir, performa model dievaluasi menggunakan beberapa metrik, yaitu accuracy, precision, recall, F1-score, serta confusion matrix untuk mengukur kualitas hasil klasifikasi.

RESULTS AND DISCUSSION

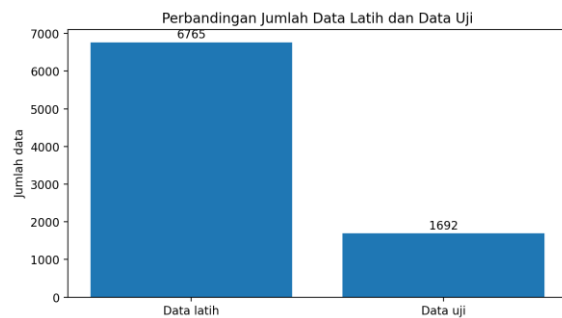
Setelah fitur terbentuk dan data sentimen tersedia, dataset dibagi menjadi data latih dan data uji. Pembagian ini bertujuan untuk melatih model pada sebagian data, kemudian menguji performanya pada data yang belum pernah dilihat sebelumnya.

Pembagian dilakukan menggunakan rasio 80:20 dengan stratifikasi, sehingga proporsi kelas sentimen pada data latih dan data uji tetap sebanding. Pada tahap klasifikasi, data yang benar-benar digunakan berjumlah 8.457 data setelah seleksi atribut yang diperlukan.

Table 1 Hasil Pembagian Data

Jenis Data	Jumlah Data	Persentase
Data latih	6.765	80%
Data uji	1.692	20%
Total	8.457	100%

Berdasarkan Tabel 1 sebagian besar data digunakan untuk melatih model, sedangkan sebagian kecil digunakan untuk pengujian. Proporsi ini umum digunakan dalam machine learning karena memberikan cukup data bagi model untuk belajar sekaligus tetap menyediakan data uji yang representatif.



Gambar 1 Perbandingan Data Latih dan Data Uji

Gambar 1 memperlihatkan bahwa jumlah data latih lebih besar dibandingkan data uji. Kondisi ini penting agar model memiliki cukup banyak contoh untuk mempelajari pola sentimen, sementara data uji tetap memadai untuk menilai kemampuan generalisasi model.

Klasifikasi Menggunakan Naive Bayes Tahap klasifikasi menggunakan algoritma Multinomial Naive Bayes yang dipilih karena efektif untuk klasifikasi teks berdimensi tinggi. Penelitian ini menggunakan dua skenario, yaitu model baseline dan model improved.

Model baseline dibangun menggunakan CountVectorizer dengan parameter default, sedangkan model improved menggunakan TF-IDF dengan n-gram (1,2) serta nilai alpha 0,5 untuk smoothing. Perbedaan konfigurasi ini dirancang untuk melihat pengaruh peningkatan fitur dan parameter terhadap performa model.

Table 2 Perbandingan Konfigurasi Model

Komponen	Baseline Model	Improved Model
Ekstraksi fitur	CountVectorizer	TF-IDF
N-gram	Unigram	Unigram + Bigram (1,2)
Parameter alpha	Default	0,5
Tujuan	Model awal	Peningkatan performa

Berdasarkan Tabel 1 model improved diharapkan dapat menangkap konteks kata dengan lebih baik karena penggunaan bigram serta pemberian bobot TF-IDF. Sementara itu, parameter alpha membantu mencegah probabilitas nol pada kata yang jarang muncul.

Evaluasi Model

Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil evaluasi membandingkan performa model baseline dan model improved untuk melihat apakah strategi peningkatan yang diterapkan benar-benar memberikan dampak positif.

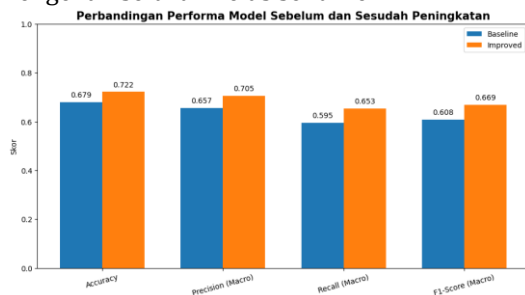
Secara umum, model improved menunjukkan performa yang lebih baik pada seluruh metrik

utama. Hal ini mengonfirmasi bahwa TF-IDF, penggunaan n-gram, dan tuning parameter Naive Bayes mampu meningkatkan kualitas klasifikasi sentimen.

Table 3 Hasil Evaluasi Model Baseline dan Improved

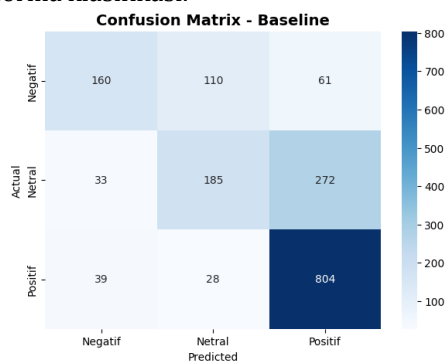
Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
Baseline	0,679	0,657	0,595	0,608
Improved	0,722	0,705	0,653	0,669

Berdasarkan Tabel 3 seluruh metrik pada model improved mengalami peningkatan dibandingkan model baseline. Accuracy meningkat dari 0,679 menjadi 0,722; precision macro meningkat dari 0,657 menjadi 0,705; recall macro meningkat dari 0,595 menjadi 0,653; dan F1-score macro meningkat dari 0,608 menjadi 0,669. Peningkatan paling besar terlihat pada F1-score macro, yang menunjukkan bahwa model improved lebih seimbang dalam mengenali seluruh kelas sentimen.



Gambar 2 Perbandingan Performa Model Sebelum dan Sesudah Peningkatan

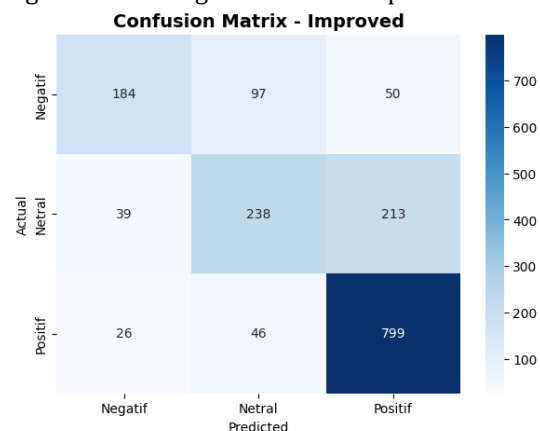
Gambar 2 menegaskan bahwa model improved unggul pada semua metrik evaluasi. Kenaikan ini menunjukkan bahwa perbaikan pada representasi fitur dan parameter model memang memberikan kontribusi nyata terhadap performa klasifikasi.



Gambar 3 Confusion Matrix Model Baseline

Pada Gambar 3 terlihat bahwa model baseline masih cukup sering salah

mengklasifikasikan kelas netral ke kelas positif. Kesalahan ini menunjukkan bahwa model baseline belum sepenuhnya mampu membedakan ulasan yang bersifat ambigu atau tidak eksplisit.



Gambar 4 Confusion Matrix Model Improved

Gambar 4 memperlihatkan perbaikan pada model improved, terutama pada kelas netral dan negatif. Pada model improved, jumlah prediksi benar meningkat dan kesalahan klasifikasi ke kelas positif berkurang. Hal ini menunjukkan bahwa strategi peningkatan model berhasil memperbaiki kemampuan generalisasi model.

Secara keseluruhan, hasil evaluasi membuktikan bahwa model improved lebih layak digunakan untuk klasifikasi sentimen ulasan Alfagift karena memberikan hasil yang lebih tinggi dan lebih seimbang pada seluruh kelas sentimen. Dengan demikian, kontribusi "peningkatan model" pada penelitian ini terlihat secara eksplisit pada hasil akhir.

CONCLUSION

Berdasarkan hasil penelitian yang telah dilakukan terhadap ulasan aplikasi Alfagift, dapat disimpulkan bahwa penerapan preprocessing teks, ekstraksi fitur TF-IDF, dan optimasi model Naive Bayes mampu meningkatkan kualitas klasifikasi sentimen secara signifikan. Hasil eksperimen menunjukkan bahwa model improved memiliki performa lebih baik dibandingkan model baseline, baik dari sisi akurasi maupun keseimbangan prediksi antar kelas.

Penerapan preprocessing teks secara komprehensif, yang meliputi cleaning, case folding, normalisasi kata, tokenizing, stopword removal, dan stemming, berhasil mengurangi noise pada data ulasan sehingga teks menjadi lebih representatif untuk proses klasifikasi. Selain itu, penggunaan TF-IDF dengan n-gram membantu model menangkap informasi kata yang lebih penting dibandingkan pendekatan frekuensi sederhana, sehingga kualitas fitur yang digunakan dalam klasifikasi menjadi lebih baik.

Peningkatan kinerja model Naive Bayes terbukti cukup signifikan setelah dilakukan optimasi. Model baseline menghasilkan accuracy sebesar 0,679, precision macro 0,657, recall macro 0,595, dan F1-score macro 0,608. Setelah dilakukan peningkatan dengan TF-IDF, n-gram, dan tuning parameter alpha, nilai tersebut meningkat menjadi accuracy 0,722, precision macro 0,705, recall macro 0,653, dan F1-score macro 0,669. Hasil ini menunjukkan bahwa model improved lebih unggul dalam mengklasifikasikan sentimen ulasan Alfagift dibandingkan model dasar.

Pola distribusi sentimen menunjukkan bahwa ulasan aplikasi Alfagift didominasi oleh sentimen positif sebanyak 4.313 data, diikuti sentimen netral sebanyak 2.468 data, dan sentimen negatif sebanyak 1.673 data. Sementara itu, kata-kata dominan yang muncul pada hasil analisis menggambarkan topik utama ulasan pengguna, terutama berkaitan dengan promo, belanja, aplikasi, dan pengiriman. Temuan ini menunjukkan bahwa model improved tidak hanya mampu mengklasifikasikan sentimen dengan lebih baik, tetapi juga memberikan gambaran yang lebih jelas mengenai karakteristik ulasan pengguna Alfagift.

REFERENCE

- [1] B. Z. Ramadhan, R. I. Adam, and I. Maulana, "Analisis Sentimen Ulasan pada Aplikasi E-Commerce dengan Menggunakan Algoritma Naive Bayes," *Journal of Applied Informatics and Computing*, vol. 6, no. 2, pp. 220–225, Dec. 2022, doi: 10.30871/JAIC.V6I2.4725.
- [2] A. Ernawati, A. O. Sari, S. N. Sofyan, M. Iqbal, and R. F. W. Wijaya, "Implementasi Algoritma Naive Bayes dalam Menganalisis Sentimen Review Pengguna Tokopedia pada Produk Kesehatan," *Bulletin of Information Technology (BIT)*, vol. 4, no. 4, pp. 533–543, Dec. 2023, doi: 10.47065/bit.v4i4.1090.
- [3] Z. A. Nurdiyansa and B. Berlilana, "Sentiment Analysis of Reviews on Lazada Apps using Naive Bayes Algorithm," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 1, pp. 594–601, Jan. 2024, doi: 10.30865/MIB.V8I1.7255.
- [4] I. Ali, D. Rakhmawati, and M. Wijaya, Anugerah Bagus, "Analyzing the Sentiment of the 2024 Election Sirekap Application Using Naive Bayes Algorithm," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 8, no. 3, pp. 1503–1511, Jul. 2024, doi: 10.30865/MIB.V8I3.7773.
- [5] Wilonotomo, B. Mulyawan, M. Ryanindityo, M. A. Syahrin, and F. Y. Triana, "Text Mining Algorithm Naive Bayes Classifier to Improve Quality Sentiment Analysis Passport Mobile Application," *Journal of Social and Political Sciences*, vol. 7, no. 1, Mar. 2024, doi: 10.31014/AIOR.1991.07.01.475.
- [6] W. Purbaratri *et al.*, "Sentiment Analysis of e-Government Service Using the Naive Bayes Algorithm," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 23, no. 2, pp. 441–452, Mar. 2024, doi: 10.30812/matrik.v23i2.3272.