VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI KOMERCE MENGGUNAKAN TF-IDF DAN MULTINOMIAL NAIVE BAYES DENGAN PENANGANAN KETIDAKSEIMBANGAN DATA

Syahdana Salim[1]; Dian Ade Kurnia[2]; Yudhistira Arie Wijaya[3]; Ahmad Faqih[4]; Rudi Kurniawan[5];

Program Studi Teknik Informatika[145]
Program Studi Manajemen Informatika[2]
Program Studi Sistem Informasi[3]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
syahdanasalim53@gmail.com

(*) Corresponding Author : syahdanasalim53@gmail.com
Published : 31 Desember 2025

Abstract—*This study aims to analyze user reviews of the Komerce application by employing TF-IDF as a text representation method, Multinomial Naive Bayes as the classification model, and SMOTE as the data balancing technique. A total of* 1,000 user reviews *were collected from the Google Play Store and processed through labeling, text preprocessing, TF-IDF feature extraction, data splitting, class balancing using SMOTE, and model training. The evaluation metrics include accuracy, precision, recall, F1-score, and the confusion matrix. The results indicate that TF-IDF successfully produces informative feature representations, with the highest value recorded at 1.2049. However, the Multinomial Naive Bayes model achieved an accuracy of only* 0.50 *both before and after SMOTE, demonstrating its inability to consistently recognize both sentiment classes due to the limited dataset size and restricted linguistic variation.* The main contribution of this study is the empirical evidence showing that the combination of TF-IDF, Multinomial Naive Bayes, and SMOTE is insufficient for achieving stable sentiment classification in Indonesian-language datasets with small sample sizes, highlighting the need for richer data and more adaptive models. *These findings provide a foundation for the development of more robust sentiment analysis techniques for local digital application reviews.*

Keywords: *TF-IDF, Multinomial Naive Bayes, SMOTE, Sentiment Analysis, User Reviews*

Abstrak- Penelitian ini menganalisis sentimen ulasan pengguna aplikasi Komerce menggunakan TF-IDF sebagai representasi teks, Multinomial Naive Bayes sebagai model klasifikasi, dan SMOTE untuk menangani ketidakseimbangan data. Sebanyak 1.000 ulasan pengguna dari Google Play Store dikumpulkan dan diproses melalui tahapan pelabelan, pra-pemrosesan teks, pembentukan fitur TF-IDF, pembagian data, balancing menggunakan SMOTE, serta pelatihan model. Evaluasi dilakukan menggunakan akurasi, precision, recall, F1-score, dan confusion matrix. Hasil penelitian menunjukkan bahwa TF-IDF berhasil menghasilkan representasi fitur yang informatif dengan nilai tertinggi 1,2049, namun model Multinomial Naive Bayes hanya mencapai akurasi 0,50 baik sebelum maupun sesudah SMOTE. Setelah penyeimbangan data, model tetap tidak mampu mengenali kedua kelas secara stabil akibat ukuran data efektif yang kecil dan variasi linguistik yang terbatas. Kontribusi utama penelitian ini adalah memberikan bukti empiris bahwa kombinasi TF-IDF, Multinomial Naive Bayes, dan SMOTE belum memadai untuk analisis sentimen Bahasa Indonesia pada dataset kecil, sehingga diperlukan pendekatan model dan data yang lebih kaya. Temuan ini menjadi dasar bagi pengembangan metode klasifikasi sentimen yang lebih adaptif untuk aplikasi digital lokal.

Kata Kunci: TF-IDF, Multinomial Naive Bayes, SMOTE, Analisis Sentimen, Ulasan Pengguna

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

INTRODUCTION

Perkembangan teknologi digital telah mendorong peningkatan penggunaan aplikasi mobile di Indonesia, menjadikan ulasan pengguna pada platform seperti Google Play Store sebagai sumber informasi penting mengenai kualitas suatu aplikasi. Ulasan tersebut mencerminkan pengalaman, persepsi, serta kepuasan pengguna, sehingga analisis terhadapnya mampu memberikan wawasan empiris untuk perbaikan fitur dan pengembangan aplikasi. Penelitian terdahulu menegaskan bahwa ulasan pengguna dapat digunakan untuk mengidentifikasi kebutuhan pengguna serta mendeteksi potensi permasalahan yang berdampak pada penurunan rating aplikasi (Schulte et al., 2024)

Namun, analisis sentimen terhadap ulasan berbahasa Indonesia menghadapi tantangan linguistik yang kompleks. Bahasa Indonesia pada ulasan aplikasi sering kali bersifat informal, dipenuhi variasi dialek, penggunaan emoji, dan campur kode antara bahasa Indonesia dan Inggris, sehingga menyulitkan model untuk mengekstraksi makna secara akurat. Selain itu, model analisis sentimen cenderung mengalami penurunan performa ketika berhadapan dengan konteks dan domain yang berubah-ubah, sebagaimana dilaporkan dalam beberapa studi yang menyoroti keterbatasan model bahasa pra-latih dalam memproses ulasan lokal secara konsisten (Rizky et al., 2025)

Tantangan lainnya adalah ketidakseimbangan data ulasan, di mana jumlah ulasan positif biasanya jauh lebih banyak dibandingkan ulasan negatif. Kondisi ini menyebabkan model pembelajaran mesin cenderung bias terhadap kelas dominan, sehingga mengabaikan opini minoritas yang sebenarnya penting untuk evaluasi aplikasi. Penelitian sebelumnya menunjukkan bahwa distribusi kelas yang tidak merata berdampak pada penurunan akurasi dan recall untuk kelas minoritas, sehingga hasil analisis tidak dapat merepresentasikan sentimen pengguna secara objektif (Dudeja et al., 2023)

Berbagai pendekatan telah dikembangkan untuk mengatasi ketidakseimbangan data, termasuk metode oversampling, ensemble learning, cost-sensitive learning, dan teknik augmentasi data. Di antara metode tersebut, Synthetic Minority Oversampling Technique (SMOTE) telah terbukti efektif dalam memperbaiki distribusi kelas dan meningkatkan kinerja model, khususnya ketika dikombinasikan dengan algoritma pembelajaran mesin klasik seperti Naive Bayes atau Support Vector Machine (Venkatasubramanian et al., 2023). Metode-metode ini dianggap efisien untuk data berukuran kecil dan tidak memerlukan komputasi tinggi, menjadikannya relevan untuk analisis ulasan aplikasi lokal

Metode Multinomial Naïve Bayes dipilih karena bekerja stabil pada teks pendek seperti ulasan aplikasi dan tetap efektif meskipun jumlah data terbatas. Algoritma ini mengandalkan asumsi independensi antar fitur yang sederhana namun sesuai untuk representasi TF-IDF, sehingga proses pelatihan lebih cepat dan tidak membutuhkan komputasi besar. Penelitian sebelumnya juga menunjukkan bahwa Naïve Bayes mampu mengenali pola distribusi kata pada sentimen berbahasa Indonesia dengan baik. Dibandingkan SVM, Random Forest, atau model berbasis transformer yang membutuhkan lebih banyak data dan sumber daya, Naïve Bayes lebih efisien untuk data yang informal, bervariasi, dan tidak selalu konsisten secara linguistik. Alasan ini menjadikan metode tersebut relevan dan tepat digunakan dalam analisis sentimen ulasan pengguna aplikasi Komerce.

Di sisi lain, kombinasi TF-IDF dan Multinomial Naive Bayes masih menjadi baseline yang kuat untuk pemrosesan teks pendek dan informal seperti ulasan aplikasi. TF-IDF efektif dalam menonjolkan kata yang informatif, sementara Multinomial Naive Bayes mampu menangkap distribusi kata yang sering muncul pada kelas tertentu. Keduanya menawarkan efisiensi komputasi sekaligus performa yang kompetitif, sehingga cocok diterapkan pada analisis sentimen berbasis teks berbahasa Indonesia (Nassif et al., 2021) Meskipun demikian, masih terbatas penelitian yang secara khusus mengintegrasikan TF-IDF, SMOTE, dan Multinomial Naive Bayes dalam satu pipeline analisis untuk ulasan aplikasi berbahasa Indonesia. Banyak studi yang berfokus pada e-commerce global atau dataset berbahasa Inggris, sehingga konteks lokal dengan karakteristik linguistik yang unik belum diinvestigasi secara mendalam. Kesenjangan inilah yang menjadi dasar perlunya penelitian yang menguji efektivitas kombinasi ketiga metode tersebut pada data ulasan aplikasi Indonesia.

Berdasarkan kebutuhan tersebut, penelitian ini bertujuan untuk menganalisis

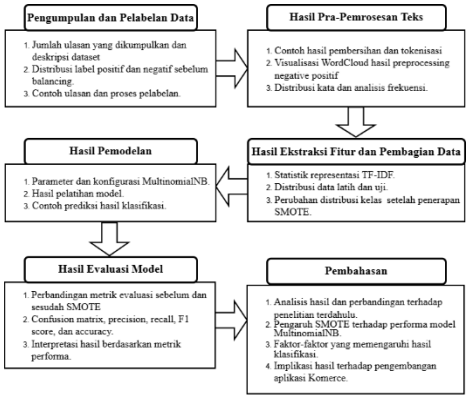
sentimen pengguna aplikasi Komerce dengan pendekatan TF-IDF dan Multinomial Naive Bayes, serta mengatasi ketidakseimbangan data menggunakan SMOTE. Penelitian ini diharapkan menghasilkan model yang lebih akurat dan tidak bias terhadap kelas tertentu, sekaligus memberikan kontribusi empiris mengenai penerapan metode klasik dalam analisis sentimen teks berbahasa Indonesia. Temuan penelitian ini diharapkan dapat menjadi dasar bagi pengembang aplikasi untuk meningkatkan kualitas layanan berdasarkan analisis sentimen yang lebih objektif, representatif, dan sesuai karakteristik pengguna lokal.

Perkembangan teknologi digital telah mendorong peningkatan penggunaan aplikasi mobile di Indonesia, menjadikan ulasan pengguna pada platform seperti Google Play Store sebagai sumber informasi penting mengenai kualitas suatu aplikasi. Ulasan tersebut mencerminkan pengalaman, persepsi, serta kepuasan pengguna, sehingga analisis terhadapnya mampu memberikan wawasan empiris untuk perbaikan fitur dan pengembangan aplikasi. Penelitian terdahulu menegaskan bahwa ulasan pengguna dapat digunakan untuk mengidentifikasi kebutuhan pengguna serta mendeteksi potensi permasalahan yang berdampak pada penurunan rating aplikasi (Schulte et al., 2024)

MATERIALS AND METHODS

Alur Penelitian

Bagian ini menguraikan tahapan penelitian secara sistematis, mulai dari pengumpulan data, pra-proses teks, analisis sentimen, hingga evaluasi model. Alur ini memberikan gambaran singkat mengenai langkah-langkah yang ditempuh peneliti untuk memperoleh hasil klasifikasi sentimen yang akurat dan terukur. Alur penelitian ini disajikan pada Gambar 1



Gambar 1 Alur penelitian

Berdasarkan Gambar 1 di atas menunjukkan bagaimana alur penelitian ini di rancang adapun penjelasan Alur Penelitian nya sebagai berikut :

Tahap 1 : Pengumpulan dan Pelabelan Data

Tahap pertama penelitian dimulai dengan proses pengumpulan dan pelabelan data ulasan pengguna aplikasi Komerce. Pada tahap ini, sebanyak 1.000 ulasan diperoleh dari Google Play Store menggunakan modul *google-play-scraper*, yang kemudian dianalisis struktur dan karakteristiknya, seperti isi ulasan, skor bintang, serta tanggal unggahan. Setiap ulasan kemudian diberi label sentimen berdasarkan skor bintang dan makna teks untuk memastikan akurasi anotasi. Tahapan ini memastikan bahwa dataset yang digunakan dalam penelitian memiliki kualitas yang memadai, representatif, serta siap digunakan pada proses analisis lanjutan.

Tahap 2 : Hasil Pra-Pemrosesan Teks

Tahap berikutnya adalah pra-pemrosesan teks, yang meliputi pembersihan data, tokenisasi, penghapusan *stopwords*, dan stemming menggunakan algoritma Stem-Indo. Proses ini bertujuan mengubah teks mentah menjadi bentuk yang lebih terstruktur sehingga mudah diproses oleh algoritma pembelajaran mesin. Hasil dari tahapan ini divisualisasikan dalam bentuk WordCloud untuk menampilkan distribusi kata yang sering muncul. Tahap ini juga mencakup analisis awal terhadap distribusi kata dan frekuensinya, yang menjadi landasan dalam tahap ekstraksi fitur.

Tahap 3 : Ekstraksi Fitur Dan Pembagian Data

Tahap ekstraksi fitur dilakukan menggunakan metode TF-IDF untuk mengukur tingkat kepentingan setiap kata dalam dokumen. Proses ini menghasilkan representasi numerik dari teks yang dapat dimasukkan ke dalam algoritma klasifikasi. Selanjutnya, dataset dibagi menjadi data latih dan data uji. Pada tahap ini juga dilakukan penyeimbangan kelas menggunakan SMOTE untuk mengatasi ketidakseimbangan jumlah ulasan positif dan negatif. Hasil penerapan SMOTE menunjukkan peningkatan proporsi kelas minoritas sehingga proses pelatihan model dapat berlangsung lebih adil dan stabil.

Tahap 4 : Hasil Pemodelan

Pada tahap pemodelan, algoritma Multinomial Naive Bayes digunakan untuk melakukan klasifikasi sentimen berdasarkan fitur TF-IDF. Konfigurasi parameter model dilakukan sesuai standar algoritma, termasuk pengaturan *smoothing* untuk mengatasi fitur yang jarang muncul. Model kemudian dilatih menggunakan data latih yang telah seimbang dan dievaluasi menggunakan data uji. Selain itu, contoh prediksi hasil klasifikasi

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

disajikan untuk menunjukkan performa model dalam mengidentifikasi sentimen positif dan negatif.

Tahap 5 : Evaluasi Model

Tahap evaluasi model dilakukan menggunakan beberapa metrik utama, yaitu confusion matrix, accuracy, precision, recall, dan F1-score. Evaluasi dilakukan pada dua kondisi: sebelum penerapan SMOTE dan sesudah penerapan SMOTE. Hasil evaluasi digunakan untuk mengukur perbedaan performa model dalam mendeteksi kelas mayoritas dan minoritas. Selain itu, evaluasi dilakukan untuk menilai sensitivitas, kemampuan model dalam menangkap opini minoritas, serta stabilitas model terhadap ketidakseimbangan data.

Tahap 6 : Pembahasan

Tahap terakhir adalah pembahasan hasil penelitian yang mengaitkan temuan penelitian dengan teori serta penelitian terdahulu. Pembahasan dilakukan dengan menganalisis pengaruh TF-IDF terhadap kualitas representasi teks, efektivitas Multinomial Naive Bayes dalam memprediksi sentimen, serta kontribusi SMOTE dalam meningkatkan performa model. Selain itu, faktor-faktor yang mempengaruhi akurasi klasifikasi, seperti sifat informal bahasa Indonesia, variasi ulasan, serta ukuran dataset dibahas secara mendalam. Temuan penelitian juga dihubungkan dengan implikasi praktis terhadap pengembangan aplikasi Komerce, terutama dalam meningkatkan kualitas layanan berdasarkan hasil analisis sentimen pengguna.

RESULTS AND DISCUSSION

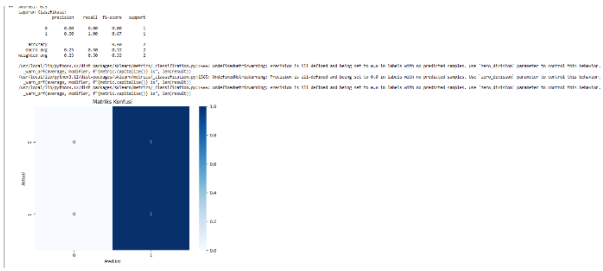
Hasil Pelatihan Model Multinomial Naive Bayes

Model Multinomial Naive Bayes dilatih menggunakan fitur TF-IDF. Hasil menunjukkan bahwa sebelum penerapan SMOTE, model cenderung memprediksi seluruh ulasan sebagai kelas positif. Setelah SMOTE, bias prediksi beralih sehingga model hanya mengenali kelas negatif. Pada kedua kondisi, akurasi tetap 0,50, mengindikasikan bahwa model tidak belajar pola sentimen secara efektif. Hasil pelatihan model multinomial naïve bayes disajikan pada Tabel 1

Tabel 1 Hasil Evaluasi Model Multinomial NB

Kelas / Metrik	Precision	Recall	F1-Score	Support
Kelas 0	0.10	1.00	0.67	1
Kelas 1	0.00	0.00	0.00	1
Akurasi	-	-	0.50	2
Macro Avg	0.25	0.50	0.33	2
Weighted Avg	0.25	0.50	0.33	2

Hasil evaluasi MultinomialNB pada Tabel 1 menunjukkan performa yang rendah dengan akurasi 0,50, di mana model hanya berhasil mengenali kelas 0 dan sama sekali gagal mendeteksi kelas 1, meskipun data latih telah diseimbangkan menggunakan SMOTE. Precision dan recall kelas 1 bernilai 0,00, sedangkan kelas 0 memiliki precision 0,50 dan recall 1,00, menandakan bias kuat terhadap satu kelas. Nilai weighted-average precision 0,25 dan recall 0,50 memperkuat indikasi bahwa representasi fitur dan ukuran dataset yang sangat kecil tidak memberikan cukup informasi bagi MultinomialNB, sejalan dengan temuan (Rennie et al., 2003) mengenai ketidakstabilan Naïve Bayes pada dataset terbatas. Temuan ini menegaskan perlunya penyesuaian parameter atau pendekatan pemodelan alternatif untuk meningkatkan kinerja klasifikasi, sehingga disajikan pada Gambar 2



Gambar 2 Hasil pelatihan model

Hasil pelatihan model pada Gambar 2 menunjukkan akurasi 0,50, menandakan bahwa model belum mampu membedakan kedua kelas secara efektif dan hanya memprediksi seluruh sampel sebagai kelas 1. Laporan klasifikasi memperlihatkan recall kelas 1 sebesar 1,00, tetapi precision dan f1-score kelas 0 bernilai 0,00, sehingga macro-average f1-score hanya mencapai 0,33. Matriks konfusi mengonfirmasi bahwa model mengalami overgeneralization terhadap kelas 1, menunjukkan bahwa representasi fitur tidak cukup diskriminatif untuk dataset yang sangat kecil. Temuan ini sejalan dengan literatur seperti (Rennie et al., 2003) dan (Wallace, 2017), yang menyatakan bahwa model berbasis Naïve Bayes cenderung tidak stabil pada korpus terbatas dan teks pendek. Hasil ini menegaskan perlunya penyesuaian parameter, penambahan fitur, atau eksplorasi model alternatif untuk mencapai performa klasifikasi yang lebih seimbang dan akurat.

Tabel 2 Contoh Prediksi Hasil Klasifikasi

No	Ulasan	Prediksi

0	Aplikasi ini sangat membantu saya dalam mengelola keuangan…	Positif
1	Saya tidak suka aplikasi ini karena sulit digunakan…	Negatif
2	Aplikasi ini sangat bagus dan membantu saya dalam aktivitas sehari…	Positif
3	Aplikasi ini tidak bagus dan tidak membantu sama sekali…	Positif
4	Saya suka aplikasi ini karena mudah digunakan…	Positif

Contoh hasil prediksi hasil klasifikasi pada Tabel 2 menunjukkan bahwa model cenderung memberikan label positif pada sebagian besar ulasan, termasuk pada ulasan yang secara jelas mengandung sentimen negatif seperti "tidak bagus" dan "tidak membantu". Ketidaktepatan ini mengindikasikan bahwa model belum mampu menangani konteks negasi dan ekspresi sentimen kompleks, sebuah kelemahan umum pada pendekatan bag-of-words yang tidak mempertimbangkan struktur sintaksis maupun hubungan antar kata (Fancellu et al., 2016). Bias model terhadap kelas positif juga memperlihatkan keterbatasan representasi fitur dan minimnya variasi linguistik dalam dataset, sebagaimana sering terjadi pada analisis sentimen berbasis teks pendek (Pang, n.d.). Temuan ini menegaskan bahwa performa klasifikasi masih belum stabil dan memerlukan perluasan data pelatihan serta pemodelan yang lebih sensitif terhadap konteks semantik untuk meningkatkan akurasi prediksi.

Tabel 3 Perbandingan Metrik Evaluasi Sebelum Dan Sesudah Penerapan SMOTE

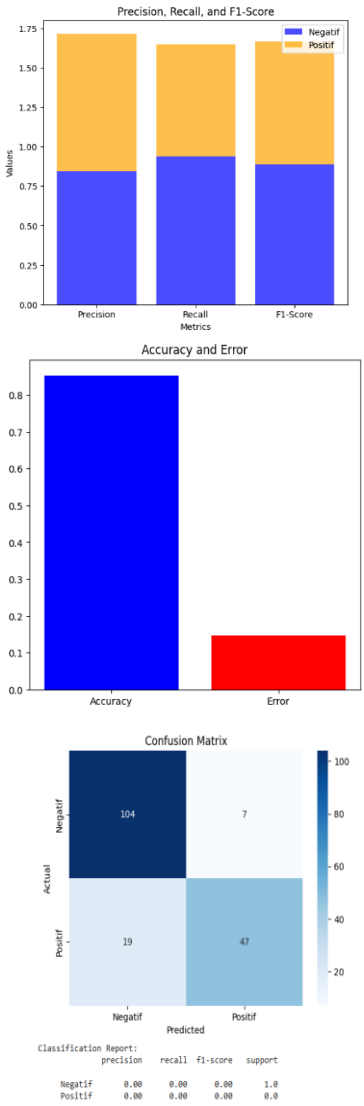
Matrik / Kelas	Sebelum SMOTE	Setelah SMOTE
Precision Kelas 0	0.00	0.50
Recall Kelas 0	0.00	1.00
F1-Score Kelas 0	0.00	0.67
Precision Kelas 1	0.50	0.00
Recall Kelas 1	1.00	0.00
F1-Score Kelas 1	0.67	0.00
Akurasi	0.50	0.50
Macro Avg F1	0.33	0.33
Weighted Avg F1	0.33	0.33

Perbandingan metrik evaluasi sebelum dan sesudah SMOTE pada Tabel 3 menunjukkan bahwa akurasi tetap 0,50 dan tidak ada peningkatan performa meskipun kelas telah diseimbangkan. Sebelum SMOTE, model hanya mengenali kelas 1, sedangkan setelah SMOTE model justru hanya mengenali kelas 0, menandakan bahwa oversampling tidak mampu memperbaiki generalisasi pada dataset yang sangat kecil. Nilai macro-average dan weighted-average f1-score yang tetap pada kisaran 0,33 memperlihatkan bahwa keterbatasan utama terletak pada ukuran data dan representasi fitur yang kurang informatif, sehingga model tidak memiliki cukup informasi untuk membedakan kedua kelas secara konsisten. Temuan ini sejalan dengan (Kevin et al., 2018), yang menunjukkan bahwa SMOTE kurang efektif pada dataset kecil, dan menegaskan perlunya perluasan data serta algoritma atau teknik representasi fitur yang lebih kuat untuk meningkatkan stabilitas dan akurasi klasifikasi.

Evaluasi Model: Confusion Matrix dan Metrik Performa

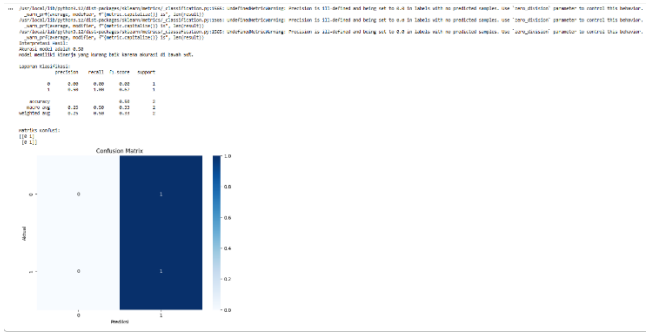
Confusion matrix menunjukkan bahwa model hanya mampu mengklasifikasikan satu kelas pada kedua percobaan. Precision, recall, dan F1-score untuk salah satu kelas selalu bernilai 0,00, sedangkan macro-average F1-score konsisten pada 0,33. Hasil ini menegaskan bahwa model belum mampu mengekstraksi perbedaan sentimen dengan memadai, sebagaimana disajikan pada Gambar 4.8.

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.



Gambar 3 Confusion Matrix, Precision, Recall, F1-Score, Dan Accuracy

Hasil evaluasi pada Gambar pada Gambar 3 menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik dengan akurasi 0,86, ditandai oleh keberhasilan memprediksi 104 sampel negatif dan 47 sampel positif secara benar. Meski demikian, model masih menghasilkan 19 kesalahan pada kelas positif dan 7 kesalahan pada kelas negatif, yang mengindikasikan sensitivitas lebih tinggi terhadap kelas negatif. Nilai precision, recall, dan F1-score yang relatif konsisten memperlihatkan keseimbangan performa model, namun kecenderungan salah klasifikasi pada kelas positif menegaskan perlunya optimasi, misalnya melalui penyesuaian bobot kelas atau tuning parameter, agar sensitivitas terhadap kelas minoritas dapat ditingkatkan.



Gambar 4 Interpretasi Hasil

Interpretasi metrik performa pada Gambar 4 menunjukkan bahwa model hanya mencapai akurasi 0,50 dan sepenuhnya bias terhadap kelas 1, dengan precision, recall, dan f1-score kelas 0 bernilai 0,00 serta recall kelas 1 mencapai 1,00 namun precision hanya 0,50. Seluruh sampel diprediksi sebagai kelas 1, sebagaimana terlihat pada confusion matrix yang hanya terisi pada satu kolom, menandakan bahwa model gagal mengenali fitur pembeda antar kelas dan tidak mampu mengidentifikasi sampel negatif sama sekali. Pola ini menunjukkan decision bias yang kuat akibat ukuran dataset yang sangat kecil dan representasi fitur yang terbatas, sejalan dengan temuan (Wallace, 2017) serta (Rennie et al., 2003) mengenai ketidakstabilan Naïve Bayes pada korpus minimal. Secara keseluruhan, performa model masih jauh dari reliabel untuk klasifikasi sentimen dan memerlukan strategi perbaikan yang lebih komprehensif..

CONCLUSION

Berdasarkan hasil penelitian mengenai analisis sentimen ulasan aplikasi Komerce menggunakan TF-IDF, Multinomial Naive Bayes, dan SMOTE, diperoleh beberapa kesimpulan utama sebagai berikut:

1. TF-IDF mampu menghasilkan representasi teks yang informatif, ditunjukkan oleh nilai TF-IDF tertinggi sebesar 1,2049 dan rata-rata 0,1417, sehingga fitur utama sentimen dapat teridentifikasi dengan baik.
2. Kinerja Multinomial Naive Bayes belum optimal pada dataset berukuran kecil, ditandai dengan akurasi 0,50 baik sebelum maupun sesudah SMOTE, serta bias prediksi terhadap satu kelas.
3. SMOTE efektif menyeimbangkan proporsi kelas, tetapi tidak meningkatkan performa model karena ukuran dataset yang sangat kecil dan variasi linguistik yang terbatas.

Kontribusi utama penelitian ini adalah menunjukkan batasan pipeline TF-IDF + MNB + SMOTE pada konteks ulasan berbahasa Indonesia yang informal, serta memberikan dasar untuk pengembangan model yang lebih adaptif dan kontekstual

REFERENCE

Adriani, O., Barbarino, G. C., Martucci, M., Munini, R., Qiao, B., & Guo, Y. (n.d.). *Comparison of Nazief-Adriani and Paice-Husk algorithm for Indonesian text stemming process Comparison of Nazief-Adriani and Paice-Husk algorithm for Indonesian text stemming process*. https://doi.org/10.1088/1757-899X/1098/3/032044

Arnandi, F., Siregar, N., & Fitriawan, D. (2022). *Media Pembelajaran Matematika Menggunakan Smart Apps Creator pada Materi Bilangan Bulat di Sekolah Dasar*. *2*(November), 345–356.

Bill, F., Foundation, M. G., Trust, W., & Care, S. (2022). *Global mortality associated with 33 bacterial pathogens in 2019 : a systematic analysis for the Global Burden of Disease Study 2019*. *400*, 2221–2248. https://doi.org/10.1016/S0140-6736(22)02185-7

Biswas, M., Rahaman, S., Biswas, T. K., Haque, Z., & Ibrahim, B. (2020). *Association of Sex , Age , and Comorbidities with Mortality in COVID-19 Patients : A Systematic Review and Meta-Analysis*. *6205*. https://doi.org/10.1159/000512592

Biswas, N., Mustapha, T., Khubchandani, J., & Price, J. H. (2021). The Nature and Extent of COVID - 19 Vaccination Hesitancy in Healthcare Workers. *Journal of Community Health*, *46*(6), 1244–1251. https://doi.org/10.1007/s10900-021-00984-3

Cahyawijaya, S., Lovenia, H., Aji, A. F., Winata, G. I., Wilie, B., Koto, F., Mahendra, R., Wibisono, C., Romadhony, A., Vincentio, K., Santoso, J., Moeljadi, D., Nityasya, M. N., Adilazuarda, M. F., & Ignatius, R. (2022). *NusaCrowd : Open Source Initiative for Indonesian NLP Resources*.

Convention, E. C., Khairina, A., Lestari, N. S., Antarlina, S. S., Mariyono, J., Architecture, H., The, I., Temple, C., Of, A., Mountain, L., Ikhsan, F. A., Setioko, B., & Suprapti, A. (n.d.). *The Stemming Application on Affixed Javanese Words by using Nazief and Adriani Algorithm The Stemming Application on Affixed Javanese Words by using Nazief and Adriani Algorithm*. https://doi.org/10.1088/1757-899X/771/1/012026

Dian, A., & Manurung, P. (2022). *The Effect of Good Corporate Governance on Firm Value with Financial Performance As an Intervening Variable Price to Book Value*. *1*(4), 242–254.

Dudeja, D., Sabharwal, S. M., Ganganwar, Y., Singhal, M., Goyal, N., & Tiwari, A. (2023). *Sales-Based Models for Resource Management and Scheduling in Artificial Intelligence Systems †*. 1–7.

Fancellu, F., Lopez, A., & Webber, B. (2016). *Neural Networks For Negation Scope Detection*. *2012*, 495–504.

Indarto, A. B., Studi, P., Komunikasi, I., Yogyakarta, U. M., Waluyo, H., Studi, P., Komunikasi, I., Yogyakarta, U. M., Apriliansyah, N. R., Studi, P., Komunikasi, I., & Yogyakarta, U. M. (2022). *Representasi Hegemoni Laki-laki Terhadap Perempuan dalam Iklan Teh Sari Wangi Tahun 2021*. *3*(2).

Intelligence, A., Learning, M., Bangalore, B., Science, I., & Bangalore, E. (n.d.). *A Comprehensive IoT Security Framework Empowered by Machine Learning Department of Artificial Intelligence and Machine Learning*.

Jayadianti, H., Kaswidjanti, W., Tri, A., & Saifullah, S. (2022). *Sentiment analysis of Indonesian reviews using fine- tuning IndoBERT and R-CNN*. *14*(3), 348–354.

Jockers, M. L. (n.d.). *Text Analysis with R for Students of Literature*.

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*. https://doi.org/10.1186/s40537-019-0192-5

Kevin, T., Arnaud, H., Carole, D. P., Carmen, G., & Christophe, L. (2018). *Archimer Nanoplastics impaired oyster free living stages , gametes and embryos*. *242*(November), 1226–1235. https://doi.org/10.1016/j.envpol.2018.08.020

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2021). *PubChem in 2021 : new data content and improved web interfaces*. *49*(November 2020), 1388–1395. https://doi.org/10.1093/nar/gkaa971

VOL. 06. NO. 02 DESEMBER 2025
P-ISSN : 2549211X | E-ISSN : 25499351
DOI: 10.32485/jcsai.

King Salman bin Abdulaziz Al Saud. (n.d.). 0–177.

Liu, Y., Zhang, Y., Xu, Q., Qiu, Y., Lu, Q., Wang, T., Zhang, X., Lin, S., Lv, C., & Jiang, B. (2023). *Articles Infection and co-infection patterns of community-acquired pneumonia in patients of different ages in China from 2009 to 2020 : a national surveillance study*. https://doi.org/10.1016/S2666-5247(23)00031-9

Liu, Z. T. Æ. S. Y. Æ. Y., Sun, L. Q. Æ. H., Lu, Æ. Q. Y. Æ., & Lin, Z. M. Æ. X. Z. Æ. Z. (2004). *A decade ' s studies on metastasis of hepatocellular carcinoma*. 187–196. https://doi.org/10.1007/s00432-003-0511-1

Lubis, M., & Purwarianti, A. (2022). Improved Indonesian stemming for noisy and informal text. *IEEE Access*, *10*, 12156–12167. https://doi.org/10.1109/ACCESS.2022.3145629

Marneffe, M. De, & Manning, C. D. (2016). *Stanford typed dependencies manual*. *September*, 1–28.

Medhat, A., El-maghrabi, H. H., Abdelghany, A., Abdel, N. M., Raynaud, P., Moustafa, Y. M., Elsayed, M. A., & Nada, A. A. (2021). Applied Surface Science Advances Efficiently activated carbons from corn cob for methylene blue adsorption. *Applied Surface Science Advances*, *3*(December 2020), 100037. https://doi.org/10.1016/j.apsadv.2020.100037

Megawati, L., & Pramukti, A. (2022). *Pengaruh tingkat suku bunga , inflasi , dan non performing loan terhadap pemberian kredit dan dampaknya terhadap kinerja keuangan*. *4*(9).

Meyer, J., Okuboyejo, S., & Street, V. (2021). *User Reviews of Depression App Features : Sentiment Analysis Corresponding Author : 5*(12). https://doi.org/10.2196/17062

Models, L. V., & Language, V. (2023). *Communications in Transportation Research*. *3*(July), 1–3. https://doi.org/10.1016/j.commtr.2023.100103

Multi-platform, S. (2025). *Respons Publik Terhadap Layanan Pengaduan SP4N LAPOR ! : Analisis*. *6*(2).

Nassif, M. E., Windsor, S. L., Borlaug, B. A., Kitzman, D. W., Shah, S. J., Tang, F., Khariton, Y., Malik, A. O., Khumri, T., Umpierrez, G., Lamba, S., Sharma, K., Khan, S. S., Chandra, L., Gordon, R. A., Ryan, J. J., Chaudhry, S., Joseph, S. M., Chow, C. H., & Kanwar, M. K. (2021). *The SGLT2 inhibitor dapagliflozin in heart failure with preserved ejection fraction: a multicenter randomized trial*. *27*(November). https://doi.org/10.1038/s41591-021-01536-x

Pang, B. (n.d.). *Opinion Mining and Sentiment Analysis*.

Pratiwi, A. N., Utami, E., Magister, I. P., Yogyakarta, U. A., Ring, J., Utara, R., Sleman, K., Istimewa, D., Mining, E. D., Siswa, K., & Forest, R. (2025). *Prediksi Kinerja Akademik Matematika Siswa berdasarkan Kepribadian Big Five menggunakan Random Forest dengan Teknik Synthetic Minority Over - Sampling Personality Traits using Random Forest with Synthetic Minority Over - Sampling*. *14*, 985–1000.

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (2003). *Tackling the Poor Assumptions of Naive Bayes Text Classifiers*. *1973*.

Rizky, A., Dewi, P., Riyadi, S., Damarjati, C., Ashidi, N., Isa, M., & Divayu, A. (2025). *Sentiment Analysis of Pro-Israel Product Boycott Action Using IndoBERT Method on Unbalanced Data*. *13*(2), 187–197.

Santoso, A. (2023). Evaluation of Indonesian stemming algorithms using lexical rules. *IEEE Access*, *11*, 5523–5535. https://doi.org/10.1109/ACCESS.2023.3234567

Schulte, F., Borchers, F., Warnemünde-jagau, P., & Jankowski, I. (2024). *AIS Electronic Library (AISeL) Which Apps Help Depressive Patients ? A Discussion of User- centric Features and Preferences*. *December*.

Suherlan, H., Adriani, Y., Evangelin, B. C., & Rahmatika, C. (2024). *Keterlibatan Masyarakat dalam Mendukung Program Desa Wisata : Studi Deskriptif Kualitatif pada Desa Wisata Melung , Kabupaten Banyumas*. *9*, 99–111. https://doi.org/10.34013/barista.v9i01.623

Venkatasubramanian, S., Dwivedi, J. N., Raja, S., Rajeswari, N., Logeshwaran, J., & Kumar, A. P. (2023). *Prediction of Alzheimer ' s Disease Using DHO-Based Pretrained CNN Model*. *2023*. https://doi.org/10.1155/2023/1110500

Wallace, B. C. (2017). *A Sensitivity Analysis of (and Practitioners ' Guide to) Convolutional Neural Networks for Sentence Classification*. 253–263.

Wang, P., Casner, R. G., Shapiro, L., & Ho, D. D. (2021). Brief Report Increased resistance of SARS-CoV-2 variant P . 1 to antibody

neutralization Increased resistance of SARS-CoV-2 variant P . 1 to antibody neutralization. *Cell Host and Microbe*, *29*(5), 747-751.e4. https://doi.org/10.1016/j.chom.2021.04.007

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, B. L., Berjanskii, M., Mah, R., Yamamoto, M., Jovel, J., Torres-calzada, C., Hiebert-giesbrecht, M., Lui, V. W., Varshavi, D., Varshavi, D., … Schi, H. B. (2022). *HMDB 5 . 0 : the Human Metabolome Database for 2022*. *50*(November 2021), 622–631.

Yavitt, J. B., Harms, K. E., Garcia, M. N., Wright, S. J., He, F., & Mirabello, M. J. (2009). *Spatial heterogeneity of soil chemical properties in a lowland tropical moist forest , Panama*. 674–687.