KARYA TULIS ILMIAH_AMALIA ROSSA ANNJANI.pdf

by - -

Submission date: 20-Jan-2026 10:39PM (UTC+0900)
Submission ID: 2860060116
File name: KARYA_TULIS_ILMIAH_AMALIA_ROSSA_ANNJANI.pdf (457.81K)
Word count: 4174
Character count: 25884

Application Of Machine Learning Algorithm To Prediction User Satisfaction Sentiment On Weverse

[Penerapan Algoritma Machine Learning Untuk Memprediksi Sentimen Kepuasan Pengguna Weverse]

Amalia Rossa Annjani[1)], Uce Indahyanti[2)*]

[1)]Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
[2)]Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
*Email Penulis Korespondensi: uceindahyanti@umsida.ac.id

Abstract. The rapid development of digital technology has increased the use of fan community platforms such as Weverse, generating user reviews as indicators of application service satisfaction. However, Google Play Store reviews are unstructured and have imbalanced sentiment class distributions, making sentiment classification challenging. This study aims to compare the performance of Decision Tree and Naïve Bayes algorithms in classifying Weverse user review sentiment. The methodology includes text preprocessing, semi-automatic labeling, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), and handling class imbalance using Random Oversampling (ROS) on training data. Model evaluation was conducted using a Confusion Matrix with data split ratios of 80:20 and 70:30. The results show that the Naïve Bayes algorithm with an 80:20 split achieved the best performance, with an accuracy of 86.81% and an F1-score of 83.88%. The application of ROS improved sentiment classification balance..

Keywords - Sentiment Analysis; Decision Tree; Naive Bayes; Random Oversampling (ROS); Weverse Review

Abstrak. Perkembangan teknologi digital mendorong meningkatnya penggunaan platform komunitas penggemar seperti Weverse yang menghasilkan ulasan pengguna sebagai indikator kepuasan layanan aplikasi. Namun, ulasan pada Google Play Store bersifat tidak terstruktur dan memiliki distribusi kelas sentimen yang tidak seimbang sehingga menyulitkan proses klasifikasi. Penelitian ini bertujuan membandingkan kinerja algoritma Decision Tree dan Naïve Bayes dalam mengklasifikasikan sentimen ulasan pengguna Weverse. Metode penelitian meliputi preprocessing teks, pelabelan semi-otomatis, ekstraksi fitur menggunakan Term Frequency–Inverse Document Frequency (TF-IDF), serta penanganan ketidakseimbangan kelas dengan Random Oversampling (ROS) pada data latih. Evaluasi model dilakukan menggunakan Confusion Matrix dengan pembagian data 80:20 dan 70:30. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes dengan rasio 80:20 menghasilkan kinerja terbaik dengan akurasi 86,81% dan F1-Score 83,88%. Penerapan ROS terbukti meningkatkan keseimbangan klasifikasi sentimen..

Kata Kunci - Analisis Sentimen; Decision Tree; Naïve Bayes; Random Oversampling (ROS); Ulasan Weverse

I. PENDAHULUAN

Perkembangan teknologi digital telah mengubah cara pengguna berinteraksi dan mengakses hiburan melalui platform daring. Weverse, sebagai platform komunitas penggemar yang dikembangkan oleh HYBE Corporation, menyediakan berbagai fitur interaksi antara artis dan penggemar melalui konten eksklusif serta forum diskusi. Pada tahun 2024, Weverse mencatat peningkatan pengguna global sebesar 19% dengan total unduhan lebih dari 150 juta. Tingginya jumlah pengguna tersebut berdampak pada meningkatnya volume ulasan yang mencerminkan pengalaman serta tingkat kepuasan pengguna terhadap layanan aplikasi[1].

Ulasan pengguna pada Google Play Store menjadi sumber informasi penting dalam menilai persepsi dan tingkat kepuasan pengguna terhadap aplikasi. Ulasan tersebut mencakup berbagai aspek, seperti antarmuka, performa, stabilitas, hingga fitur aplikasi. Namun, ulasan bersifat tidak terstruktur sehingga sulit dianalisis secara langsung tanpa proses pengolahan teks. Oleh karena itu, diperlukan metode analisis sentimen untuk mengidentifikasi kecenderungan sentimen positif atau negatif dari ulasan pengguna[2].

Berbagai penelitian sebelumnya telah memanfaatkan algoritma machine learning untuk analisis sentimen pada platform digital, seperti Twitter, YouTube, dan e-commerce. Meskipun demikian, penelitian yang berfokus pada platform komunitas penggemar seperti Weverse masih terbatas. Selain itu, belum terdapat penelitian yang secara khusus membandingkan kinerja algoritma Decision Tree dan Naïve Bayes dalam mengklasifikasikan kepuasan pengguna Weverse. Oleh karena itu, kedua algoritma tersebut diusulkan untuk diterapkan pada analisis sentimen ulasan pengguna Weverse

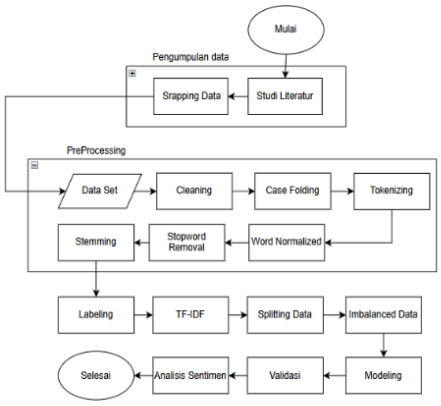
Berdasarkan kesenjangan tersebut, penelitian ini dilakukan untuk membandingkan kinerja algoritma Decision Tree dan Naïve Bayes dalam mengklasifikasikan sentimen ulasan pengguna Weverse. Teknik Term Frequency–Inverse Document Frequency (TF-IDF) digunakan untuk mengubah teks ulasan menjadi representasi numerik[3]. Mengingat distribusi kelas sentimen yang tidak seimbang, teknik Random Oversampling (ROS) diterapkan pada data latih untuk menyeimbangkan kelas tanpa memengaruhi data uji. Evaluasi kinerja model dilakukan menggunakan confusion matrix guna memperoleh hasil pengukuran yang objektif dan akurat. Penelitian ini diharapkan dapat memberikan gambaran mengenai persepsi pengguna terhadap aplikasi Weverse serta menjadi referensi bagi pengembangan fitur dan penelitian selanjutnya[4].

II. METODE

Penelitian ini bertujuan memprediksi sentimen kepuasan pengguna Weverse dengan menggunakan algoritma machine learning pada ulasan Google Play Store. Algoritma yang pertama adalah Decision Tree, yang membentuk aturan keputusan berbasis perhitungan Entropy dan Information Gain[2]. Algoritma kedua adalah Naive Bayes, yang mengklasifikasikan teks berdasarkan probabilitas kata dengan asumsi independensi antar kata[5]. Penelitian terdahulu menunjukkan bahwa Naive Bayes memberikan akurasi lebih tinggi pada komentar YouTube setelah tahap preprocessing[6]. Studi lain menemukan bahwa Decision Tree lebih tepat digunakan untuk ulasan layanan kurir meskipun memiliki potensi bias data[3]. Lalu penelitian lain yang menggunakan Naive Bayes dan Decision Tree untuk menganalisis sentimen cryptocurrency di Twitter, tetapi terbatas pada ukuran dataset dan akurasi rendah[7]. Mengingat studi sebelumnya belum banyak meneliti platform komunitas digital seperti Weverse, penelitian ini melakukan analisis komparatif kedua algoritma untuk mengklasifikasikan sentimen ulasan pengguna. Penelitian ini diharapkan memberikan kontribusi baru pada analisis sentimen aplikasi digital.

II.1 Tahap Penelitian

Penelitian ini dilakukan melalui beberapa tahapan sebagaimana ditunjukkan pada Gambar 1. Rangkaian proses dimulai dari Studi Literatur, Scrapping Data, Preprocessing Data, Labelling, ekstraksi fitur menggunakan TF-IDF, Splitting Data, Imbalanced Data, Modeling, Validasi, hingga tahap Analisis. Seluruh prosedur diimplementasikan menggunakan bahasa Python, yang dikenal sebagai bahasa pemrograman interpretatif dengan sintaks sederhana, mudah dibaca, dan dapat dijalankan pada berbagai platform. Pengolahan data dan eksekusi program dilakukan melalui Google Colab, yaitu platform komputasi berbasis cloud yang memungkinkan pengguna menjalankan kode Python langsung melalui browser[8].



Gambar 1. Rancangan Penelitian

II.2 Studi Literatur

Penelitian ini bertujuan memperdalam pemahaman analisis sentimen dalam machine learning, dengan fokus pada algoritma Decision Tree dan Naive Bayes, menggunakan referensi dari jumal ilmiah dan sumber digital.

II.3 Scraping Data

Data ulasan pengguna Weverse dikumpulkan dari Google Play Store menggunakan Google Play Scraper di Google Colab, sehingga diperoleh 10.000 data kotor dari periode Januari 2020 hingga Januari 2025. Jumlah tersebut dianggap memadai karena tidak ada standar baku ukuran dataset pada analisis sentimen, dan penelitian sebelumnya juga menggunakan data yang lebih sedikit, seperti 2.000 ulasan aplikasi Ajaib[9]. Dan juga seperti 2.534 ulasan di X mengenai Tweet masyarakat pada isu Indonesia Gelap[10]. Meskipun masih mengandung berbagai bahasa, slang, dan karakter tidak relevan, data ini menjadi dasar kuat sebelum memasuki tahap preprocessing. Atribut yang digunakan dalam penelitian mencakup Ulasan (content), Rating (score), Date (at), dan Id (review_id). Contoh hasil scrapping dapat dilihat pada Tabel 1.

Tabel 1.Dataset Hasil Scraping

Id	Rating	Ulasan	Date
459fd1f5-be0a-450f9b46-aec0c8af8e91	5	Suka bangett	2025-01-28
59dd2e28-7b2f4417-a908-09b511ea185b	4	Bagus kenapa harus update	2025-01-28
bd5f1a97-9942- 424b-b673-8aa011b80168	3	lumayan	2025-01-28
381a0c26-5733- 4f22-a0c5-430678bec1af	2	Terlalu cepat update	2025-01-27
8d4ee6d3-dda5- 4bef-9f93-267543f8647c	1	Burik	2025-01-27

II.4 Preprocessing Data

Preprocessing data dilakukan untuk menghilangkan noise dan menyiapkan teks agar dapat digunakan pada tahap pemodelan, sehingga data menjadi lebih bersih dan konsisten[11]. Tahapan preprocessing meliputi cleaning untuk menghapus simbol, emoji, dan karakter tidak relevan[2], case folding untuk mengubah seluruh huruf menjadi huruf kecil[12], tokenizing untuk memecah teks menjadi token kata[13], word normalization untuk mengubah kata tidak baku menjadi kata baku[14], stopword removal untuk menghilangkan kata umum yang kurang berpengaruh terhadap sentimen[15], serta stemming menggunakan library Sastrawi untuk mereduksi kata ke bentuk dasarnya[16].

II.5 Labelling

Untuk memastikan kualitas label pada dataset pelatihan dan pengujian, setiap ulasan diberikan label sentimen (Positif atau Negatif) menggunakan metode pelabelan semi-otomatis berbasis aturan. Penentuan label diawali dengan memanfaatkan nilai rating, di mana rating 1–2 langsung diberi label negatif dan rating 4–5 diberi label positif. Ulasan dengan rating 3 dianggap ambigu sehingga diproses lebih lanjut menggunakan Kamus Leksikon Sentimen. Skor leksikon diperoleh dari penjumlahan bobot kata, di mana skor > 0 diklasifikasikan sebagai positif, sedangkan skor ≤ 0 diklasifikasikan sebagai negatif. Dengan cara ini, seluruh ulasan berhasil dikategorikan ke dalam dua kelas sentimen tanpa menyisakan kelas netral, sesuai ruang lingkup penelitian[17].

II.6 TF-IDF

TF-IDF (Term Frequency–Inverse Document Frequency) digunakan untuk mengubah teks ulasan menjadi representasi numerik berdasarkan tingkat kemunculan kata dan seberapa penting kata tersebut dalam keseluruhan dokumen. Metode ini membantu model mengenali kata-kata yang memiliki bobot kuat dalam membedakan sentimen antar ulasan[18].

II.7 Splitting & Imbalanced Data

Data berlabel dibagi menjadi dua subset yaitu data pelatihan dan data pengujian. Pembagian ini dilakukan dengan menggunakan rasio 80:20 dan 70:30 untuk mengevaluasi kinerja model di berbagai distribusi data[19]. Dikarenakan dataset memiliki ketidakseimbangan kelas (imbalanced data), di mana jumlah ulasan positif lebih banyak dibandingkan ulasan negatif. Untuk mengatasi hal tersebut, diterapkan teknik Random Oversampling (ROS) pada data train dengan cara menambah sampel pada kelas negatif sehingga distribusi kelas menjadi seimbang selama proses pelatihan. Sementara itu, data test tetap dibiarkan dalam bentuk aslinya agar distribusi datanya tetap mencerminkan kondisi nyata dan untuk mencegah terjadinya data leakage. Dengan pendekatan ini, model dapat mempelajari pola secara lebih merata namun tetap dievaluasi menggunakan data dengan distribusi asli[4].

II.8 Modeling & Validasi

Model machine learning dilatih pada data pelatihan menggunakan algoritma Decision Tree dan Naive Bayes. Kemudian, prediksi model dibandingkan dengan label aktual untuk menilai kemampuannya dalam mengklasifikasikan sentimen secara akurat. Kemudian dilanjutkan dengan Model klasifikasi yang telah dilatih divalidasi menggunakan metrik evaluasi yang berasal dari Confusion Matrix untuk mengukur kinerja sentimen secara akurat dan menyeluruh. Presisi (Precision) digunakan untuk menilai keandalan model dalam memprediksi kelas Positif, yaitu seberapa banyak prediksi Positif yang benar-benar tepat. Recall, dalam konteks penelitian ini, menjadi metrik penting karena dataset bersifat imbalanced. Recall digunakan untuk melihat kemampuan model dalam menangkap seluruh data pada kelas minoritas, terutama kelas Negatif yang rentan terlewat. Setelah dilakukan proses Random Oversampling (ROS), nilai recall meningkat secara signifikan, menandakan bahwa model menjadi lebih mampu mengenali ulasan Negatif. Sementara itu, F1- Score tetap menjadi metrik gabungan yang memberikan gambaran kinerja keseluruhan secara seimbang dan mendukung analisis sentimen pada ulasan pengguna aplikasi Weverse[20].

III. Hasil dan Pembahasan

III.1 Hasil Preprocessing

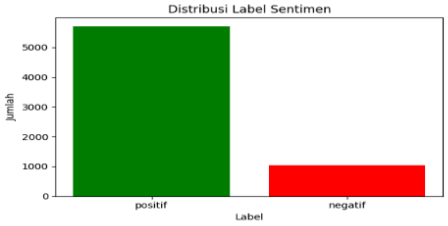
Pada tahap ini dilakukan pembersihan data sehingga dataset berkurang dari 10.000 menjadi 6.742 data bersih. Pengurangan jumlah data ini disebabkan oleh proses penghapusan ulasan duplikat, ulasan kosong, ulasan yang hanya berisi simbol atau karakter tidak relevan, serta entri yang tidak dapat diproses akibat struktur teks yang rusak. Setelah proses pembersihan ini, dilakukan pelabelan sentimen menggunakan pendekatan berbasis aturan. Contoh hasil preprocessing dataset beserta hasil pelabelan sentimen disajikan pada Tabel 2.

Tabel 2. Dataset Setelah Preprocessing dan Labelling

Ulasan Awal	Ulasan Akhir	Rating	Label	Date
Suka bangett	suka sangat	5	positif	2025-01-28
Bagus kenapa harus update	bagus	4	positif	2025-01-28
lumayan	cukup	3	positif	2025-01-28

Terlalu cepat update	terlalu cepat	2	negatif	2025-01-27
Burik	Jelek	1	negatif	2025-01-27

Pada Gambar 2. Distribusi label sentimen pada dataset hasil pelabelan menunjukkan adanya ketidakseimbangan kelas. Jumlah data dengan sentimen positif lebih dominan dibandingkan sentimen negatif, yaitu masing-masing sebanyak 5.706 data positif dan 1.035 data negatif. Kondisi ini menunjukkan bahwa sebagian besar ulasan pengguna Weverse cenderung bersentimen positif dan perlu diperhatikan pada tahap pemodelan selanjutnya.



Gambar 2. Distribusi Label Sentimen

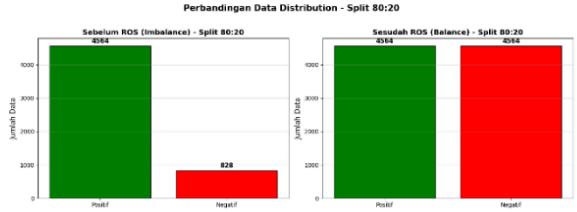
III.2 Imbalanced Data (ROS)

Untuk memberikan gambaran yang lebih jelas mengenai hasil proses penyeimbangan data, Tabel 3 menyajikan distribusi data train sebelum dan sesudah penerapan Random Oversampling (ROS). Penyajian ini bertujuan menunjukkan perubahan proporsi kelas yang terjadi setelah penyeimbangan, sehingga proses pelatihan model berlangsung dengan representasi kelas yang lebih merata. Adapun data test tetap ditampilkan dalam kondisi aslinya untuk menjaga validitas evaluasi.

Tabel 3. Distribusi dan Penyeimbangan Data Train dan Test

Rasio Split	Kelas Sentimen	Jumlah Data Awal (Imbalance)	Jumlah Data ROS (Balance)	Data Test (Asli)
80:20	Positif (Mayoritas)	4.564	4.564	1.142
	Negatif (Minoritas)	828	4.564	207
	Total Data Train	5.392	9.128	1.349
70:30	Positif (Mayoritas)	3.994	3.994	1.712
	Negatif (Minoritas)	724	3.994	311
	Total Data Train	4.718	7.988	2.023

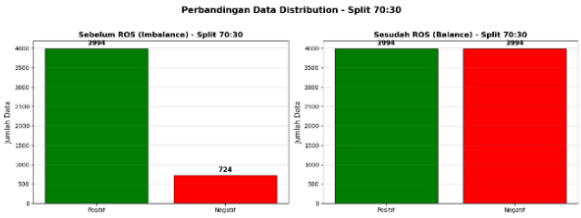
Pada Gambar 3, terlihat bahwa pada rasio pembagian data 80:20, distribusi data train sebelum penerapan Random Oversampling (ROS) masih menunjukkan ketidakseimbangan kelas, di mana jumlah data sentimen positif lebih dominan dibandingkan sentimen negatif. Setelah dilakukan penyeimbangan menggunakan ROS, jumlah data pada kelas positif dan negatif menjadi sama, sehingga distribusi data train menjadi lebih merata untuk proses pelatihan model.



Gambar 3. Perbandingan Data Distribusi Rasio 80:20

Selanjutnya, pada Gambar 4, distribusi data train dengan rasio 70:30 menunjukkan pola yang serupa. Sebelum penerapan Random Oversampling (ROS), data sentimen positif mendominasi dibandingkan data sentimen negatif.

Setelah dilakukan penyeimbangan, jumlah data pada kedua kelas menjadi seimbang, yang diharapkan dapat mengurangi bias model terhadap kelas mayoritas pada tahap pelatihan.



Gambar 4. Perbandingan Data Distribusi Rasio 70:30

III.3 Hasil Pengujian Algoritma

Pengujian model dilakukan terhadap delapan skenario yang berbeda. Skenario ini termasuk perbandingan dua algoritma klasifikasi (Decision Tree dan Naive Bayes), dua rasio pembagian data (80:20 dan 70:30), dan dua kondisi data latih (Balance dan Imbalance). Tujuan pengujian adalah untuk membandingkan kinerja kedua algoritma sekaligus mengevaluasi pengaruh penyeimbangan data menggunakan Random Oversampling (ROS) terhadap kinerja model, khususnya kemampuan untuk mengidentifikasi kelas minoritas. Hasil perbandingan untuk kedua rasio pembagian data disajikan pada Tabel 4 dan Tabel 5, yang menunjukkan bahwa kondisi balance (ROS) berdampak lebih besar pada kinerja model daripada kondisi imbalance (tanpa ROS)

Tabel 4. Hasil Komparasi Rasio Split 80:20

Algoritma	Kondisi Train	Akurasi	Precision	Recall	F1-Score
Decision Tree	Imbalance	85.47%	85.36%	85.47%	85.41%
Decision Tree	Balance (ROS)	84.14%	85.36%	84.14%	84.67%
Naïve Bayes	Imbalance	86.81%	85.10%	86.81%	83.88%
Naïve Bayes	Balance (ROS)	81.10%	88.03%	81.10%	83.15%

Tabel 5. Hasil Komparasi Rasio Split 70:30

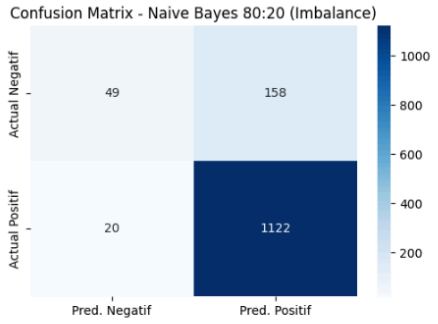
Algoritma	Kondisi Train	Akurasi	Precision	Recall	F1-Score
Decision Tree	Imbalance	83.84%	83.65%	83.84%	83.74%
Decision Tree	Balance (ROS)	82.80%	84.08%	82.80%	83.73%
Naïve Bayes	Imbalance	86.51%	84.99%	86.51%	82.95%
Naïve Bayes	Balance (ROS)	80.43%	86.98%	80.43%	82.47%

III.4 Analisis Komparasi Kinerja Model

Hasil pengujian menunjukkan bahwa algoritma Naive Bayes, yang memiliki rasio split 80:20 pada kondisi Imbalance, memiliki akurasi tertinggi. Namun, dia memiliki kinerja yang lebih buruk pada kelas negatif, seperti yang ditunjukkan oleh nilai Recall yang rendah, yang membuat sebagian besar ulasan negatif tidak dapat ditemukan. Random Oversampling (ROS) telah terbukti meningkatkan kemampuan model untuk mengidentifikasi kelas minoritas secara signifikan, mengingat betapa pentingnya mengidentifikasi ulasan negatif bagi pengembang untuk mengevaluasi masalah layanan. Namun, ini menyebabkan penurunan akurasi dan F1-Score Weighted secara keseluruhan. Seperti yang ditunjukkan oleh fenomena ini, ada trade-off yaitu akurasi kelas mayoritas dikurangi untuk meningkatkan Recall pada kelas minoritas, yang membuat model lebih seimbang dan sensitif terhadap ulasan negatif[21]. Kondisi ini dapat dilihat pada Tabel 6 analisis Confusion Matrix, di mana nilai False Positive yang tinggi pada model tanpa ROS menunjukkan bahwa kelas mayoritas bias. Dengan menggunakan ROS, distribusi data latih menjadi lebih seimbang, meningkatkan kemampuan deteksi ulasan negatif dan mendukung tujuan penelitian dalam mengevaluasi sentimen pengguna aplikasi Weverse.

Tabel 6. Komparasi Kinerja Kelas Negatif

Algoritma	Split Rasio	Kondisi Train	Recall	F1-Score
Decision Tree	80:20	Imbalance	51.69%	52.20%
		Balance (ROS)	57.49%	52.65%
	70:30	Imbalance	45.98%	46.66%
		Balance (ROS)	53.05%	48.67%
Naïve Bayes	80:20	Imbalance	23.67%	35.51%
		Balance (ROS)	81.64%	57.00%
	70:30	Imbalance	18.97%	30.18%
		Balance (ROS)	76.85%	54.69%



Gambar 5. Confusion Matrix Model Optimal

VII. SIMPULAN

Berdasarkan hasil pengujian dan analisis komparasi yang dilakukan terhadap klasifikasi sentimen ulasan menggunakan algoritma Decision Tree (DT) dan Naïve Bayes (NB), didapatkan bahwa konfigurasi model Naïve Bayes (NB) dengan Rasio Split 80:20 pada kondisi Imbalance adalah yang terbaik secara keseluruhan. Konfigurasi ini mencapai akurasi tertinggi sebesar 86.81% dan F1-Score Weighted sebesar 83.88%. Peran Rasio Split 80:20 mendukung keunggulan ini. Namun, model ini memiliki kelemahan, yaitu nilai False Positive (FP) yang tinggi, dan menunjukkan bias model terhadap kelas Mayoritas. Dalam penelitian ini, penggunaan Random Oversampling (ROS) menghasilkan trade-off, yaitu akurasi keseluruhan menurun, tetapi kemampuan untuk mendeteksi dan mengingat pada kelas negatif (minoritas) secara signifikan meningkat, yang menjadikan model lebih seimbang. Oleh karena itu, untuk mengevaluasi kemampuan generalisasi model, disarankan untuk melakukan pengujian algoritma klasifikasi tambahan sebagai perbandingan, mempelajari metode penyeimbangan data tambahan untuk mengatasi bias secara lebih tepat, dan menguji dataset ulasan yang lebih besar.

UCAPAN TERIMA KASIH

Penulis berterima kasih kepada Universitas Muhammadiyah Sidoarjo karena telah memberikan bantuan dan fasilitas selama penyusunan artikel ini. Ucapan terimakasih juga ditujukan kepada dosen pembimbing karena telah memberikan bimbingan, arahan, dan saran yang sangat berharga. Selain itu, penulis mengucapkan terima kasih kepada semua orang yang telah membantu, mendukung, dan motivasi, baik secara langsung maupun tidak langsung, dalam proses menyelesaikan artikel ini.

REFERENSI

[1] A. Z. Tofani, "WWeverse Sebagai Sarana Komunikasi Fans Dengan Idol(Studi Pada Interaksi Seventeen Dan Carat)," vol. 1, pp. 349–357, 2023.

[2] E. F. Laili et al., "Komparasi Algoritma Decision Tree Dan Support Vector Machine (Svm) Dalam Klasifikasi Serangan Jantung," J. Sist. Inf. dan Inform., vol. 8, no. 1, pp. 67–76, 2025.

[3] D. N. I. Huda, C. Prianto, and R. M. Awangga, "Dellavianti Nishfi Ilmiah Huda Analisis Sentimen Perbandingan Layanan Jasa Pengiriman Kurir Pada Ulasan Play Store Menggunakan Metode Random Forest dan Descision Tree," J. Ilm. Inform., vol. 11, no. 2, pp. 150–158, 2023.

[4] Z. Abidin, T. Suratno, and M. F. Putri, "PENERAPAN RANDOM OVERSAMPLING DAN PRINCIPAL COMPONENT ANALYSIS UNTUK MENINGKATKAN AKURASI PREDIKSI KEBANGKRUTAN PERUSAHAAN DI INDONESIA DENGAN MODEL MACHINE LEARNING," vol. 12, no. 5, 2025.

[5] M. I. Abas, R. Lamusu, W. E. Pranata, S. Syahrial, and I. Ibrahim, "Penerapan Algoritma Naive Bayes Untuk Sistem Klasifikasi Status Gizi Bayi Balita," vol. 5, no. 5, pp. 1249–1260, 2025.

[6] B. Franko, N. Wilyanto, H. Irsyad, U. Multi, and D. Palembang, "Analisis Sentimen Terhadap Naturalisasi Pemain pada Youtube Menggunakan Decision Tree dan Naive Bayes Sentiment Analysis of Player Naturalization on Youtube Using Decision Trees and Naive Bayes," vol. 03, no. September, pp. 8–16, 2024.

[7] A. Syahrul, A. I. Purnamasari, and I. Ali, "Analisis Sentimen Twitter Terhadap Cryptocurrency P (C | X) =," JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 2, pp. 1–8, 2024.

[8] M. I. Prayugah, U. Indahyanti, and N. Ariyanti, "ANALISIS SENTIMEN PUBLIK PADA PEMERINTAH DALAM SERANGAN RANSOMWARE DENGAN PENDEKATAN SMOTE," vol. 8, no. 2, pp. 333–343, 2024.

[9] N. H. Nanda Dwi Kurniawan, Praditya Rendi Ferdian, "Analisis Sentimen Algoritma Naïve Bayes, Support Vector Machine, dan Random Forest Pada Ulasan Aplikasi Ajaib," J. Nas. Teknol. dan Sist. Inf., vol. 01, pp. 87–97, 2025.

[10] S. Ramadhani, Taufik Hidayat, "ANALISIS SENTIMEN TWEET MASYARAKAT PADA ISU INDONESIA GELAP DI MEDIA SOSIAL X MENGGUNAKAN ALGORITMA NAÏVE BAYESDAN METODE COUNTVECTORIZER," JATI(Jurnal Mhs. Tek. Inform., vol. 9, no. 6, pp. 9664–9670, 2025.

[11] G. Kanugrahan, V. Hafizh, C. Putra, and Y. Ramdhani, "Analisis Sentimen Aplikasi Gojek Menggunakan SVM , Random Forest dan Decision Tree," vol. 6, no. 2, 2024.

[12] S. A. Lubis, Rahma, "Klasifikasi Sentiment Analysis TerhadapUsulan KB Vasektomi Syarat Penerima BANSOS dengan MetodeNaive Bayes," JUNRAL KOMPUTER SISTEM INFORMASI KOMPUTER, 2025. [Online]. Available: https://ejurnal.lkpkaryaprima.id/index.php/juktisi/article/view/631/408.

[13] A. S. Ngabdul Basidt, Eko Supriyadi, "Perbandingan Algoritma Klasifikasi dalam Analisis Sentimen Opini Masyarakat tentang Kenaikan Harga Bbm," Joined J. (Journal Informatics Educ., vol. 6, no. 2, p. 219, 2024.

[14] L. Rhomaningtias, A. Khairunisa, S. Shella, M. Wara, and K. M. Hindrayani, "ANALISIS SENTIMEN ULASAN APLIKASI SMILE INDONESIA MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)," HOAQ J. Teknol. Inf., vol. 16, no. c, pp. 79–91, 2025.

[15] F. F. T. Andy Kho, "Analisis Sentimen Pengguna Aplikasi STEAM dengan Algoritma Naive Bayes," J. Multidiscip. Res. Dev., vol. 7, no. 6, pp. 4620–4627, 2025.

[16] T. E. Puput Amelia Effendi, "ANALISIS SENTIMEN ULASAN APLIKASI GAMEHAY DAY MENGGUNAKAN ALGORITMA RANDOM FOREST," J. Inform. dan Tek. elektro Terap., vol. 13, no. 3, pp. 1–8, 2025.

[17] D. N. Febianty and M. Rahardi, "A Sentiment Analysis of Public Perception Toward Pets in Public Spaces Using Logistic Regression and Word Embedding," J. Appl. Informatics Comput., vol. 9, no. 4, pp. 1846–1851, 2025.

[18] N. A. V. Andi Nur Halim, Rudiman Rudiman, "Analisis Sentimen Opini Publik Terhadap Peristiwa Bitcoin Halving Pada Data Teks Twitter Menggunakan Metode Naïve Bayes Dan Pembobotan Fitur TF-IDF," vol. 4, no. 3, pp. 2823–2831, 2025.

[19] W. M. B. Najib Nurdiansyah, Farhan Sulis Febriyan, Zanuar Gesit Dian Amanta, Dicky Arya Saputra, "Mental Health Analysis to Prevent Mental Disorders in Students Using The K-Nearest Neighbor (K-NN) Algorithm and Random ForestAlgorithmAnalisis Kesehatan Mental untuk Mencegah Gangguan Mental pada MahasiswaMenggunakan Algoritma K-Nearest Neighbor(K-NN) da," nstitut Ris. dan Publ. Indones. Indones. J. Mach. Learn. Comput. Sci., vol. 5, no. January, pp. 1–9, 2025.

[20] D. S. Wijaya and D. Widyaningrum, "Komparasi metode algoritma klasifikasi pada analisis sentimen komentar cyberbullying di instagram," vol. 7, pp. 466–472, 2024.

[21] H. P. Jelita and W. Saad, Muhammad Ibnu, "Penerapan Algoritma Naïve Bayes Dalam Analisis Sentimen Masyarakat Terhadap STMIK Widya Cipta Dharma," Bull. Inf. Technol., vol. 6, no. 2, pp. 148–160, 2025.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

KARYA TULIS ILMIAH_AMALIA ROSSA ANNJANI.pdf

ORIGINALITY REPORT

20%	18%	9%	18%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Exeed College Student Paper	18%
2	Heppy Nur Asavia Ginasputri, Atika Dwi Saputri, Siti Nurasriyanti Wahid, Ariska Fitriyana Ningrum, M Al Haris. "Analisis Sentimen Ulasan Pengguna BRImo Terhadap Pembaruan Fitur Aplikasi Menggunakan Naive Bayes Dengan Seleksi Fitur Chi-Square", Jurnal Statistika dan Komputasi, 2025 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		