PENERAPAN ALGORITMA MACHINE LEARNING UNTUK MEMPREDIKSI SENTIMEN KEPUASAN PENGGUNA *WEVERSE*

Amalia Rossa Annjani, Uce Indahyanti*, Azmuri Wahyu Azinar, Rohman Dijaya
Informatika, Universitas Muhammadiyah Sidoarjo
Jl. Raya Gelam No.250, Pagerwaja, Gelam, Kec. Candi, Kabupaten Sidoarjo, Jawa Timur, Indonesia
uceindahyanti@umsida.ac.id

ABSTRAK

Perkembangan teknologi digital mendorong meningkatnya penggunaan platform komunitas penggemar seperti *Weverse*, yang menghasilkan ulasan pengguna sebagai indikator kepuasan terhadap layanan aplikasi. Namun, ulasan pada Google Play Store bersifat tidak terstruktur dan memiliki distribusi kelas sentimen yang tidak seimbang, sehingga menyulitkan proses klasifikasi sentimen. Permasalahan utama dalam penelitian ini adalah ketidakseimbangan data sentimen serta pemilihan algoritma klasifikasi yang mampu memberikan kinerja optimal. Penelitian ini bertujuan untuk 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 menggunakan *Random Oversampling* (ROS) pada data latih. Evaluasi model dilakukan menggunakan *Confusion Matrix* dengan rasio pembagian data sebesar 80:20 dan 70:30. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* dengan rasio split 80:20 menghasilkan kinerja terbaik dengan nilai akurasi sebesar 86,81% dan F1-Score sebesar 83,88%. Penerapan *Random Oversampling* (*ROS*) terbukti meningkatkan *recall* pada kelas negatif sehingga menghasilkan model klasifikasi yang lebih seimbang.

Kata kunci : *Analisis Sentimen, Decision Tree, Naïve Bayes, Random Oversampling (ROS), Ulasan Weverse,*

1. 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]

2. TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka dan Teori

Dalam rangka memprediksi sentimen kepuasan pengguna *Weverse*, penelitian ini membandingkan kinerja dua algoritma *Machine Learning*, yaitu *Decision Tree* dan *Naïve Bayes*. *Decision Tree* bekerja dengan memodelkan keputusan sebagai struktur pohon, di mana pemilihan atribut terbaik untuk percabangan didasarkan pada perhitungan Gain tertinggi dari nilai *Entropy*, menghasilkan serangkaian aturan sederhana yang mengarah ke label sentimen[2]. Sementara itu, *Naïve Bayes* adalah metode probabilistik yang mengklasifikasikan ulasan

berdasarkan *Teorema Bayes* dan mengukur Probabilitas Posterior suatu kelas sentimen (H) setelah melihat kata dalam ulasan (X), dengan asumsi kuat bahwa semua kata saling independen[5]. Kinerja kedua model ini dievaluasi secara objektif menggunakan metrik *Confusion Matrix*, mencakup Akurasi, Presisi, *Recall*, dan F1-Score. F1-Score secara khusus digunakan untuk menentukan keseimbangan kinerja model terhadap setiap kelas sentimen, menjadi dasar utama untuk memilih model yang paling optimal.

2.2. Penelitian Terdahulu

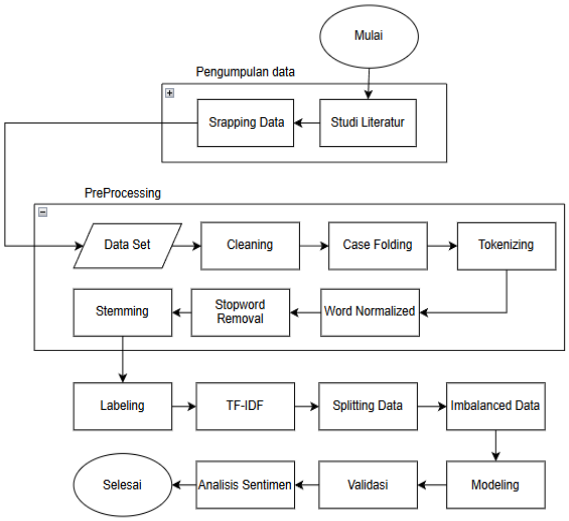
Perbandingan antara algoritma *Decision Tree* dan *Naive Bayes* dalam menganalisis sentimen dari komentar *YouTube* terkait naturalisasi pemain telah dilakukan, yang menemukan bahwa *Naive Bayes* menghasilkan akurasi yang lebih tinggi setelah tahap *preprocessing*[6]. Studi lain menganalisis sentimen terhadap *cryptocurrency* di Twitter menggunakan *Naïve Bayes* dan *Decision Tree* untuk mengklasifikasikan sentimen, namun penelitian tersebut memiliki kelemahan pada ukuran dataset yang terbatas dan tingkat akurasi yang rendah[7]. Lebih lanjut, penelitian yang mengevaluasi analisis sentimen layanan jasa pengiriman kurir berdasarkan ulasan di Play Store dengan membandingkan performa *Random Forest* dan *Decision Tree*, yang menunjukkan keunggulan pada penggunaan dua algoritma untuk analisis yang lebih menyeluruh, tetapi dengan keterbatasan potensi bias data[3]. Secara keseluruhan, temuan-temuan ini menunjukkan bahwa algoritma *Decision Tree* dan *Naïve Bayes* merupakan dua metode yang paling sering digunakan dalam berbagai konteks analisis sentimen, mulai dari media sosial (*Twitter, YouTube*) hingga *e-commerce*. Meskipun hasil akurasi kedua algoritma bervariasi bergantung pada karakteristik dataset dan tahap *preprocessing* yang diterapkan, studi komparatif menunjukkan bahwa kedua metode ini tetap relevan dan sesuai untuk klasifikasi teks.

2.3. Analisis GAP

Meskipun analisis sentimen berbasis *machine learning* telah diterapkan secara luas di berbagai bidang seperti media sosial, *e-commerce*, dan transportasi, tetapi masih belum banyak penelitian yang berfokus secara khusus pada platform komunitas digital seperti *Weverse*. Kemudian, penelitian ini menggunakan ulasan pengguna dari Google Play Store untuk mengetahui pengalaman pengguna yang sebenarnya, berbeda dengan penelitian sebelumnya yang menggunakan data survei ataupun data dari Twitter. Selain itu, penelitian ini memperkenalkan analisis komparatif dari dua teknik klasifikasi yang berbeda yaitu *Decision Tree* (berdasarkan logika) dan *Naive Bayes* (berdasarkan probabilitas) yang sebelumnya belum pernah diterapkan bersama dalam konteks aplikasi *Weverse*. Hal ini menawarkan kontribusi baru pada bidang analisis sentimen untuk aplikasi digital.

3. METODE 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

3.1. Studi Literatur

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

3.2. 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-450f-9b46-aec0c8af8e91	5	Suka bangett	2025-01-28
59dd2e28-7b2f-4417-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

3.3. Preprocessing Data

Dataset yang telah dikumpulkan diproses untuk menghilangkan *noise* dan menyiapkan teks agar dapat digunakan pada tahap modeling. *Preprocessing* dilakukan untuk mengubah data mentah menjadi struktur yang lebih bersih, rapi, dan konsisten[11].

3.4. Cleaning

Cleaning dilakukan untuk menghapus simbol, emoji, tanda baca berlebih, dan karakter tidak relevan lainnya. Sebagai contoh, teks ulasan "baguss bbgttt□□", "Kenapa harus selalu di update???? Kesel" setelah proses cleaning menjadi "baguss bbgttt", "kenapa harus selalu di update kesel", sehingga teks lebih bersih dan siap untuk tahap pemrosesan selanjutnya. [2].

3.5. Case Folding

Case folding adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil. Langkah ini dilakukan untuk memastikan konsistensi data dan menghindari perbedaan makna akibat perbedaan penulisan huruf kapital[12].

3.6. Tokenizing

Tokenizing adalah proses memecah teks ulasan menjadi potongan kata atau token. Tahap ini memudahkan analisis lanjutan dan diperlukan sebelum proses stopword removal dilakukan[13].

3.7. Word Normalized

Word normalization adalah proses mengubah kata tidak baku, typo, atau bentuk slang menjadi kata baku sesuai kaidah bahasa Indonesia. Tahap ini dilakukan untuk memastikan setiap kata memiliki bentuk yang konsisten sehingga meminimalkan variasi kata yang tidak diperlukan dalam analisis[14].

3.8. Stopword Removal

Stopword removal adalah proses menghapus kata-kata umum yang kurang berpengaruh terhadap penentuan sentimen. Tahap ini dilakukan agar model lebih fokus pada kata bermakna penting dalam teks ulasan[15].

3.9. Stemming

Stemming adalah proses mereduksi kata menjadi bentuk dasarnya menggunakan *library* Sastrawi. Sebagai contoh, kata "menonton", "mengubah", dan "melihat" diubah menjadi "tonton", "ubah", dan "lihat", sehingga setiap kata memiliki bentuk dasar yang seragam dan sesuai dengan kaidah bahasa Indonesia. [16].

3.10. 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].

3.11. Splitting Data

Data berlabel dibagi menjadi dua s*ubset* 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[18].

3.12. 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[19].

3.13. Imbalanced Data

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].

3.14. Modeling

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[20].

3.15. Validasi

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].

4. HASIL DAN PEMBAHASAN

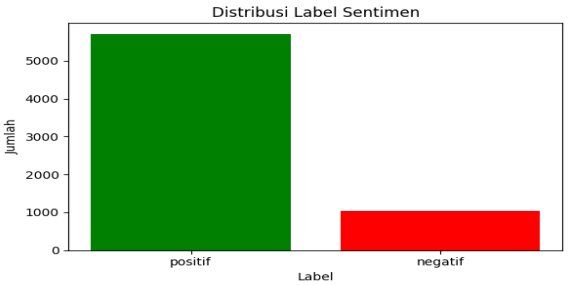
4.1. Hasil Preprocessing

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

Berdasarkan 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.



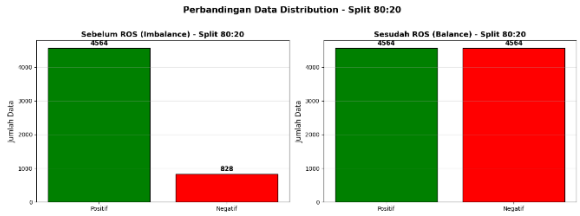
Gambar 2. Distribusi Label Sentimen

4.2. Penyeimbangan 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

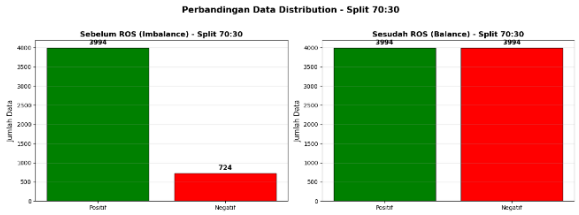
Rasio Split	Kelas Sentimen	Jumlah Data Awal (Imbalance)	Jumlah Data ROS (Balance)	Data Test (Asli)
80:20	Positif (Mayoritas)	4.564	4.564	1.142
	Negatif (Minoritas)	828	4.564	207
	Total Data Train	5.392	9.128	1.349
70:30	Positif (Mayoritas)	3.994	3.994	1.712
	Negatif (Minoritas)	724	3.994	311
	Total Data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



Gambar 3. Perbandingan Data Distribusi Rasio 80:20

Berdasarkan 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 4. Perbandingan Data Distribusi Rasio 70:30

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.

4.3. Hasil Pengujian Algoritma

Pengujian model dilakukan terhadap 8 skenario berbeda, yang mencakup perbandingan 2 algoritma klasifikasi, 2 rasio pembagian data, dan 2 kondisi data latih (*Balance* dan *Imbalance*). Tujuan utama komparasi ini adalah membandingkan kinerja *Decision Tree* dan *Naïve Bayes*, sekaligus mengevaluasi dampak penyeimbangan data latih menggunakan metode *Random Oversampling* (*ROS*) terhadap metrik kinerja, terutama F1-Score. Hasil pengujian ini disajikan secara terpisah berdasarkan rasio pembagian data.

4.4. Hasil Pengujian Rasio Split 80:20

Pengujian ini menggunakan 80% data latih dan 20% data uji. Hasil komparasi antara kondisi data *train Imbalanced* (tanpa ROS) dan *Balanced* (dengan ROS) pada rasio 80:20 disajikan pada Tabel 4.

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%

4.5. Hasil Pengujian Rasio Split 70:30

Pengujian ini menggunakan 70% data latih dan 30% data uji. Hasil komparasi antara kondisi data *train Imbalanced* (tanpa ROS) dan *Balanced* (dengan ROS) pada rasio 70:30 disajikan pada Tabel 5.

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%

4.6. Model Optimal dan Analisis Rasio Split

Berdasarkan hasil pengujian pada Tabel 4 dan Tabel 5, algoritma *Naïve Bayes* (NB) dengan rasio split 80:20 pada kondisi *imbalance* memang menunjukkan nilai Akurasi tertinggi sebesar 86,81% dan F1-Score sebesar 83.88%. Namun, performa pada kelas negatif menunjukkan kelemahan signifikan, ditunjukkan oleh nilai *Recall* negatif yang hanya mencapai 23.67%. sehingga sebagian besar keluhan pengguna tidak berhasil terdeteksi. Mengingat bahwa identifikasi ulasan negatif lebih penting bagi pengembang dalam mengevaluasi masalah layanan, maka model dengan kemampuan lebih baik dalam menangkap kelas negatif perlu diprioritaskan. Penerapan *Random Oversampling* (*ROS*) terbukti meningkatkan performa tersebut secara substansial, dengan peningkatan *Recall* kelas negatif dari 23.67% menjadi 81,64%, sehingga menghasilkan model yang lebih seimbang dan lebih selaras dengan tujuan penelitian. Oleh karena itu, pemilihan model tidak hanya didasarkan pada Akurasi tertinggi, tetapi juga pada kemampuannya dalam mengidentifikasi ulasan negatif secara memadai.

4.7. Dampak ROS dan Trade-Off Kinerja

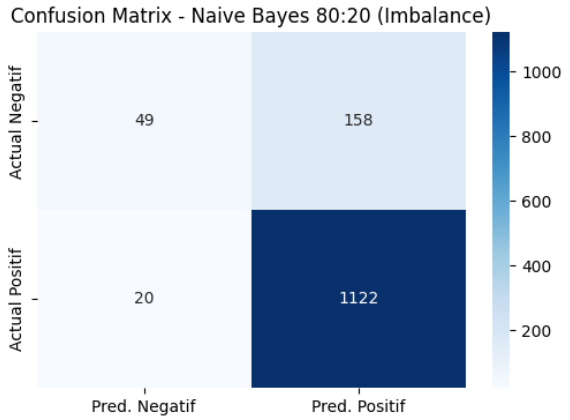
Penerapan *Random Oversampling* (*ROS*) pada data latih menyebabkan Akurasi dan F1-Score *Weighted* model menurun di semua skenario pengujian, menjadikan model *Imbalance* lebih unggul secara Akurasi keseluruhan (misalnya, *Naïve Bayes* turun dari 86.81% menjadi 81.10%). Meskipun demikian, *ROS* berhasil mencapai tujuan utamanya, yaitu mengatasi bias model terhadap kelas Mayoritas. Hal ini dibuktikan pada Tabel 6 yang menunjukkan F1-Score dan *Recall* untuk Kelas Negatif (Minoritas) meningkat secara drastis. Fenomena ini adalah sebuah *trade-off*. Akurasi (yang didominasi kelas Mayoritas) dikorbankan untuk mendapatkan *Recall* yang tinggi pada kelas Minoritas. Hasil *trade-off* ini adalah hal yang baik dan diperlukan dalam kasus data tidak seimbang, karena ini menunjukkan bahwa model telah beralih menjadi lebih seimbang dan lebih sensitif terhadap kasus kritis (negatif), meskipun penurunan Akurasi terjadi akibat peningkatan *False Positive* (FP) dari data sintetis ROS.[21]

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%

4.8. Interpretasi Metrik dan Confusion Matrix

Analisis ini berfokus pada Gambar 5 yang menunjukkan *Confusion Matrix* model terbaik yang dipilih, yaitu *Naïve Bayes* (80:20 *Imbalance*). Kekuatan model yang mendorong Akurasi 86.81% terletak pada tingginya nilai *True Positive* (TP), di mana model berhasil memprediksi sebagian besar ulasan Positif dengan tepat. Namun, *Confusion Matrix* ini juga mengungkapkan kelemahan model melalui tingginya nilai *False Positive* (FP). Angka FP yang tinggi menunjukkan bahwa sejumlah ulasan yang seharusnya Negatif, salah diklasifikasikan sebagai Positif oleh sistem. Tingginya FP ini sekaligus menjadi cerminan dari rendahnya *Recall* pada kelas Negatif, seperti yang terbukti secara numerik pada Tabel 6. Oleh karena itu, *Random Oversampling* (*ROS*) diterapkan sebagai upaya yang diperlukan untuk menyeimbangkan distribusi data latih, dengan tujuan utama meningkatkan kemampuan deteksi dan *Recall* pada kelas Minoritas yang rentan terhadap bias.



Gambar 5. Confusion Matrix Model Optimal

5. KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian dan analisis komparasi yang dilakukan terhadap klasifikasi sentimen ulasan menggunakan algoritma *Decision Tree* (DT) dan *Naïve Bayes* (NB), ditarik kesimpulan bahwa konfigurasi model *Naïve Bayes* (NB) dengan Rasio Split 80:20 pada kondisi *Imbalance* merupakan model optimal dengan kinerja terbaik secara Akurasi keseluruhan, mencapai Akurasi tertinggi sebesar 86.81% dan F1-Score *Weighted* 83.88%. Keunggulan ini didukung oleh peran Rasio Split 80:20 yang secara konsisten memberikan kinerja lebih baik. Meskipun demikian, model optimal ini memiliki kelemahan berupa tingginya nilai *False Positive* (FP) yang mencerminkan bias model terhadap kelas Mayoritas. Penerapan *Random Oversampling* (*ROS*) dalam penelitian ini menghasilkan *trade-off*, di mana Akurasi keseluruhan menurun namun berhasil meningkatkan kemampuan deteksi dan *Recall* pada Kelas Negatif (Minoritas) secara drastis, sehingga menghasilkan model yang lebih seimbang. Oleh karena itu, sebagai saran lanjutan, disarankan untuk melakukan pengujian algoritma klasifikasi lain sebagai perbandingan, mengeksplorasi teknik penyeimbangan data lain untuk mengatasi bias secara lebih tepat, dan menguji dataset ulasan yang lebih besar untuk mengukur kemampuan generalisasi model.

DAFTAR PUSTAKA

[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] 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.

[19] 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.

[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.