MODEL PREDIKSI PERSEPSI PENGGUNA DUOLINGO BERDASARKAN KOMENTAR DIGITAL

Brigide Tirenia Loresta, Uce Indahyanti*, Irwan Alnarus Alkautsar, Mochamad Alfan Rosid
Teknik Informatika, Universitas Muhammadiyah Sidoarjo
Jl. Raya Gelam No.250, Pagerwaja, Gelam, Kec. Candi, Kabupaten Sidoarjo, Jawa Timur, Indonesia
uceindahyanti@umsida.ac.id

ABSTRAK

Duolingo merupakan aplikasi pembelajaran bahasa yang banyak digunakan dan memperoleh beragam ulasan pengguna di Google Play Store. Ulasan tersebut mencerminkan persepsi pengguna terhadap kualitas dan efektivitas aplikasi. Permasalahan yang muncul adalah banyaknya data teks tidak terstruktur serta ketidakseimbangan distribusi kelas sentimen, sehingga diperlukan metode klasifikasi yang mampu menghasilkan prediksi akurat dan seimbang. Penelitian ini bertujuan membangun model prediksi persepsi pengguna Duolingo serta membandingkan kinerja algoritma Support Vector Machine (SVM) dan Decision Tree dalam mengklasifikasikan sentimen positif dan negatif. Dataset yang digunakan berjumlah 10.000 ulasan periode 2020–2025. Tahapan penelitian meliputi preprocessing, pelabelan semi-otomatis berbasis leksikon, ekstraksi fitur menggunakan TF-IDF, pembagian data 80:20 dan 70:30, serta penanganan imbalance menggunakan Random Oversampling (ROS). Evaluasi dilakukan dengan confusion matrix menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa SVM pada rasio 80:20 dengan kondisi imbalance memperoleh akurasi tertinggi sebesar 93,57%, sedangkan penerapan ROS meningkatkan recall kelas negatif menjadi 56,19%. Dengan mempertimbangkan keseimbangan kinerja, SVM dengan ROS dinilai sebagai model yang paling optimal.

Kata kunci : *Analisis Sentimen, Duolingo, Support Vector Machine (SVM), Decision Tree, TF-IDF, Random Oversampling, Machine Learning.*

1. PENDAHULUAN

Duolingo merupakan salah satu aplikasi pembelajaran bahasa paling populer di dunia dengan lebih dari 50 juta pengguna aktif harian dan sekitar 130 juta pengguna aktif bulanan pada tahun 2025, yang menunjukkan tingginya keterlibatan pengguna [1], [2]. Aplikasi ini tersedia di Google Play Store dan dapat diakses melalui perangkat seluler maupun situs web, sehingga memungkinkan pengguna belajar kapan saja dan di mana saja. *Duolingo* menyediakan berbagai pilihan bahasa, termasuk bahasa Inggris, dengan fitur interaktif seperti gambar kosakata, audio pengucapan asli, latihan terjemahan, dan permainan penyusunan kalimat, yang membuat proses belajar lebih menarik, efektif, dan fleksibel bagi mahasiswa [3].

Seiring meningkatnya penggunaan *Duolingo*, pengguna juga aktif memberikan ulasan dan penilaian di Google Play Store. Ulasan tersebut dapat dimanfaatkan untuk mengetahui persepsi dan tanggapan pengguna terhadap kualitas serta efektivitas aplikasi. Analisis sentimen digunakan untuk mengidentifikasi kecenderungan opini pengguna dengan mengolah teks ulasan menjadi kategori positif atau negatif. Dengan demikian, analisis sentimen berperan penting dalam mengevaluasi kualitas dan efektivitas aplikasi pembelajaran seperti *Duolingo* [4].

Menurut penelitian-penelitian terbaru, ulasan aplikasi banyak berupa teks tidak terstruktur sehingga memerlukan pendekatan otomatis melalui machine learning dan Natural Language Processing untuk mengklasifikasikan sentimen secara efisien dan akurat. Dalam studi analisis sentimen ulasan aplikasi, teknik otomatis seperti klasifikasi teks dan pengolahan bahasa alami terbukti efektif dalam mengidentifikasi opini pengguna sebagai positif atau negatif serta mendukung evaluasi kualitas aplikasi secara sistematis.[5]

Dalam analisis sentimen, pemilihan algoritma klasifikasi menjadi faktor krusial yang memengaruhi performa model. Diperlukan algoritma klasifikasi yang efektif untuk mengolah data teks antara lain *Support Vector Machine* (SVM) dan *Decission Tree*. *Decision Tree* bekerja dengan membentuk struktur pohon keputusan yang mudah dipahami, sedangkan *Support Vector Machine* (SVM) mencari batas pemisah terbaik antar kelas sehingga mampu menghasilkan akurasi yang tinggi. Kedua algoritma ini banyak digunakan karena performanya yang baik dalam permasalahan klasifikasi [6].

Penelitian terdahulu menunjukkan efektivitas kedua algoritma tersebut. Alifiana et al. (2023) menggunakan 1.500 data ulasan Google Play Store dan memperoleh akurasi 81% dengan Naïve Bayes serta 85% dengan SVM, sehingga SVM menunjukkan performa yang lebih baik [7]. Penelitian oleh Rokhman et al. (2021) menggunakan 3.312 data dan memperoleh akurasi 90,20% untuk SVM serta 89,80% untuk *Decision Tree*, yang kembali menunjukkan keunggulan SVM [8].

Berdasarkan hal tersebut, penelitian berjudul *"Model Prediksi Persepsi Pengguna Duolingo Berdasarkan Komentar Digital"* menggunakan metode SVM dan *Decision Tree* bertujuan untuk mengklasifikasikan sentimen pengguna ke dalam kategori positif dan negatif berdasarkan komentar digital yang diperoleh dari *kaggle*. Selain itu,

penelitian ini juga bertujuan untuk membandingkan performa algoritma Support Vector Machine dan Decision Tree dalam proses analisis sentimen, karena penelitian sebelumnya menunjukkan perbedaan performa kedua algoritma tersebut dalam klasifikasi teks ulasan, di mana SVM cenderung menghasilkan akurasi yang lebih tinggi dibandingkan Decision Tree.[8]

Penelitian ini dilakukan dengan menerapkan beberapa tahapan sistematis, yaitu pra-pemrosesan teks dan ekstraksi fitur dilakukan menggunakan metode TF-IDF yang merupakan langkah inti dalam mempersiapkan teks untuk klasifikasi. Karena di penelitian ini distribusi kelas tidak seimbang, maka dilakukan metode untuk menyeimbangkan data menggunakan Random Oversampling[9]. Model klasifikasi kemudian dibangun menggunakan algoritma Support Vector Machine dan Decision Tree. Evaluasi kinerja model menggunakan confusion matrix untuk menghitung akurasi, presisi, recall, dan F1-score.

Hasil penelitian ini diharapkan dapat menjadi masukan bagi pengembang dalam meningkatkan kualitas dan responsivitas aplikasi *Duolingo*.

2. TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka dan Teori

Dalam rangka membangun Model Prediksi Persepsi Pengguna *Duolingo* Berdasarkan Komentar Digital, penelitian ini berfokus pada perbandingan kinerja dua algoritma *machine learning*, yaitu *Support Vector Machine* (SVM) dan *Decision Tree*, dalam mengklasifikasikan persepsi pengguna yang terekam dalam komentar digital. Algoritma SVM bekerja dengan memetakan data teks ke dalam ruang berdimensi tinggi dan menentukan *hyperplane* optimal yang mampu memisahkan kelas persepsi pengguna secara maksimal. Sementara itu, *Decision Tree* melakukan proses klasifikasi dengan membangun struktur pohon keputusan melalui pemilihan atribut terbaik pada setiap simpul berdasarkan kriteria tertentu[6]. Evaluasi kinerja kedua algoritma dalam penelitian ini dilakukan secara objektif menggunakan *confusion matrix* yang mencakup metrik akurasi, presisi, recall, dan F1-score. Nilai F1-score digunakan untuk menilai keseimbangan kemampuan model dalam mengklasifikasikan setiap kelas persepsi pengguna, sehingga menjadi dasar penentuan metode yang paling optimal dalam memprediksi persepsi pengguna *Duolingo* berdasarkan komentar digital.

2.2. Penelitian Terdahulu

Penelitian Alifiana et al. menggunakan algoritma SVM dan *Naïve Bayes* dengan dataset 1.500 ulasan Google Play Store. Hasil penelitian menunjukkan bahwa Naïve Bayes memperoleh akurasi 81%, presisi 80%, dan recall 98%, sedangkan *Support Vector Machine* mencapai akurasi 85%, presisi 87%, dan recall 94%. Dengan demikian, SVM memiliki akurasi lebih tinggi, yaitu 4% dibandingkan *Naïve Bayes*[7]**.** Penelitian lainnya menganalisis sentimen terhadap kinerja Timnas Indonesia di era Shin Tae Yong menggunakan algoritma *Support Vector Machine* (SVM) yang dioptimasi dengan Particle Swarm Optimization (PSO). Hasil penelitian menunjukkan bahwa sebelum optimasi, model SVM mencapai akurasi 72,1%, sedangkan setelah diterapkan PSO akurasinya meningkat menjadi 75,9%[10]. Penelitian Prayugah *et al.* menganalisis sentimen publik terhadap penanganan serangan ransomware oleh pemerintah menggunakan 6.484 komentar YouTube dengan algoritma SVM, *Random Forest*, dan *Naïve Bayes* serta penerapan *SMOTE*. Hasil penelitian menunjukkan bahwa *SMOTE* meningkatkan kinerja model, dengan SVM mencapai akurasi tertinggi sebesar 96%, diikuti *Random Forest* 95% dan *Naïve Bayes* 89%, sehingga SVM dengan *SMOTE* dinilai sebagai metode paling optimal[11]. Penelitian Afandi *et al.* membandingkan algoritma *Naïve Bayes*, *Decision Tree*, dan *Random Forest* dalam memprediksi kelulusan mahasiswa menggunakan data akademik angkatan 2020–2021. Hasilnya menunjukkan bahwa *Random Forest* memperoleh akurasi tertinggi sebesar 97,50%, diikuti *Decision Tree* 96,25% dan Naïve Bayes 86,25%, sehingga Random Forest dinilai sebagai algoritma paling optimal[12]. Penelitian Burnama et al. mengevaluasi kinerja beberapa algoritma klasifikasi, yaitu Random Forest, Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Logistic Regression, dan K-Nearest Neighbors (KNN), pada analisis sentimen komentar YouTube terkait turnamen MPL Season 13. Berdasarkan hasil evaluasi kinerja model, Support Vector Machine (SVM) menunjukkan performa paling unggul di antara algoritma klasifikasi tunggal dengan tingkat akurasi sebesar 86,16% pada skenario pembagian data latih dan uji 70:30. Selanjutnya, Logistic Regression dan Random Forest juga memperlihatkan kinerja yang kompetitif dengan masing-masing akurasi sebesar 85,74% dan 85,47%. Di sisi lain, algoritma Naïve Bayes, Decision Tree, serta K-Nearest Neighbors (KNN) menghasilkan nilai akurasi yang lebih rendah sehingga performanya berada di bawah ketiga algoritma tersebut. Selanjutnya, penggabungan seluruh algoritma menggunakan metode ensemble machine learning dengan pendekatan hard voting mampu meningkatkan performa klasifikasi dengan akurasi tertinggi sebesar 86,70%, melampaui kinerja masing-masing model individual.[13]

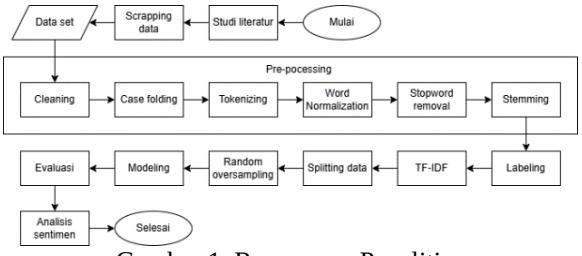
2.3. Analisis GAP

Sebagian besar penelitian mengenai analisis sentimen ulasan di Google Play Store lebih banyak fokus pada aplikasi internasional atau aplikasi yang bersifat umum. Sementara itu, penelitian yang berkaitan dengan aplikasi pembelajaran bahasa, seperti *Duolingo*, masih sangat terbatas. Selain itu, peneliti terdahuu menerapkan algoritma seperti *Decision Tree*, SVM, *Random Forest* dan belum ada

penelitian yang secara spesifik membandingkan algoritma SVM dan *Decision Tree*. Untuk itu penelitian ini dilakukan untuk membandingkan kinerja dua algoritma tersebut untuk mengetahui algoritma yang optimal serta penelitian yang dihasilkan dapat lebih relevan dan memberikan manfaat yang lebih besar bagi pengembang aplikasi *Duolingo* dalam memahami sentimen pengguna.

3. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan yang ditampilkan dalam diagram alur seperti pada Gambar 1. Proses ini di mulai dari Scrapping data, preprocessing data, labeling, TF-IDF, splitting data, modeling, evaluasi, dan juga analisis sentimen. Penelitian ini akan dilakukan menggunakan bahasa *python*. Program akan dijalankan melalui Google Colab. Google Colab merupakan platform cloud untuk menulis, menjalankan, dan membagikan kode *python* langsung melalui web browser[11]



Gambar 1. Rancangan Penelitian

3.1. Studi Literatur

Studi literatur dilakukan untuk mendapatkan berbagai referensi yang relevan dengan algoritma yang digunakan untuk penelitian. Sumber informasi yang digunakan diambil dari jurnal ilmiah dan berbagai media.

3.2. Scrapping Data

Pengumpulan data ulasan mengambil dataset dari Google Playstore. Data yang sudah diolah kemudian dipublikasikan dan di unggah ke platform *Kaggle* pada tautan berikut: https://www.kaggle.com/datasets/brigideloresta/duolingo-app-reviews-sentiment-dataset. Jumlah data ulasan yang dikumpulkan sebanyak 10.000 data dari periode januari 2020 sampai januari 2025. Dari total 10.000 data awal yang telah dikumpulkan, dilakukan proses penyaringan dengan mengambil hanya ulasan yang berbahasa Indonesia. Hasilnya, jumlah data berkurang menjadi 6.839 ulasan berbahasa indonesia. Seluruh *source code* diunggah ke repositori GitHub sebagai bentuk dokumentasi, transparansi, dan kemudahan replikasi penelitian pada tautan berikut: https://github.com/BrigideTireniaLoresta/SentimenDuolingo. Dataset ini mengambil atribut nama (userName), rating (score), dan komentar (content). Contoh hasil dari scrapping dapat dilihat pada Tabel 1.

Tabel 1. Dataset Hasil Scrapping

nama	rating	komentar
Reski Anti	5	Bagus banget sumpah
Aznii Fara	4	karena mudah dan seru
firza rahmatullah	3	lumayan sangat membantu saya
Dindaa	2	duolingo sekarang ngeselin, banyak iklannya gabisa offline pula
Ihsan Mahmudi	1	Saya beri bintang satu, karena saya sudah berlangganan tetapi sering terjadi bug dibagian test berbicara

3.3. Preprocessing Data

Setelah proses pengumpulan data, tahap selanjutnya adalah preprocessing yang difokuskan pada ulasan berbahasa Indonesia. Data ulasan aplikasi dikumpulkan dan diproses untuk menghilangkan unsur-unsur yang tidak relevan dengan analisis sentimen, misalnya karakter, angka, tanda baca, dan spasi ganda.[14]

3.4. Cleaning

Cleaning yaitu menghapus kata yang tidak relevan dan akan mencakup simbol, mention, URL, emoji, dan data kosong. Contoh kata tidak relevan yaitu “Bagus sekali Terima kasih😍❤️”, “kerenn!!!”, dan setelah mengalami proses cleaning menjadi “Bagus sekali Terimakasih”, “kerenn”.[11]

3.5. Case Folding

Pada tahapan case folding sistem merubah huruf kapital menjadi huruf kecil secara keseluruhan. Proses ini bertujuan untuk menyeragamkan bentuk kata dalam teks sehingga tidak terjadi perbedaan makna akibat perbedaan penggunaan huruf besar dan huruf kecil.[7]

3.6. Tokenizing

Tokenizing adalah memecah sebuah teks menjadi unit-unit yang lebih kecil dan bermakna agar teks tersebut dapat diproses dan dianalisis lebih lanjut oleh sistem. Proses ini membantu sistem memahami struktur teks dengan lebih baik pada tahap pengolahan data selanjutnya.[15]

3.7. Word Normalization

Word normalization yaitu mengubah kata slang menjadi kata baku, seperti mengubah "ngak" menjadi "tidak" atau "spt" menjadi "seperti" sesuai KBBI. Proses normalisasi dilakukan dengan memanfaatkan *dictionary* kata baku yang diperoleh dari platform *Kaggle*, yaitu dataset kamus slang yang tersedia pada tautan:
https://www.kaggle.com/datasets/fornigulo/kamus-slag. Kamus tersebut digunakan untuk mengubah kata-kata tidak baku menjadi bentuk baku secara sistematis dan terstruktur. Hal ini dilakukan untuk meminimalisir kata yang tidak di perlukan pada analisis.[16]

3.8. Stopword Removal

Stopword removal yaitu membuang kata yang tidak bermakna penting seperti "dan", "aku", "dari", dan lain-lain.[7]

3.9. Stemming

Stemming yaitu merubah kembali kata dalam teks menjadi kata dasar. Proses ini menggunakan library sastrawi[17]

3.10. Labelling

Proses pelabelan dilakukan secara semi-otomatis menggunakan teknik *lexicon-based labeling*, yaitu pemberian label secara otomatis berdasarkan pencocokan kata dalam ulasan dengan kamus sentimen yang dikelompokkan ke dalam kategori positif dan negatif. [18] Penentuan label awal dilakukan berdasarkan rating, di mana ulasan dengan rating 1 dan 2 dikategorikan sebagai sentimen negatif, sedangkan rating 4 dan 5 sebagai sentimen positif. ulasan dengan rating 3 ditentukan menggunakan kamus leksikon dengan menghitung akumulasi bobot kata, Jika kata positif dominan atau > 0 maka akan ditetapkan positif, dan jika kata negatif dominan atau ≤ 0 maka akan ditetapkan negatif.

3.11. Tf-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) adalah teknik yang digunakan untuk menentukan bobot pentingnya kata dalam sebuah dokumen dengan mempertimbangkan dua faktor utama: seberapa sering kata tersebut muncul dalam dokumen tertentu (Term Frequency) dan seberapa jarang kata tersebut ditemukan dalam kumpulan dokumen (Inverse Document Frequency)[19]

a. Menghitung TF (Term Frequency)

$$\mathrm{TF}_{t,d} = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Jumlah total kata dalam dokumen } d} \quad (1)$$

b. Menghitung IDF (Inverse Document Frequency)

$$\mathrm{IDF}_t = \log\left(\frac{\text{Jumlah total dokumen } (N)}{\text{Jumlah dokumen yang mengandung kata } t(DF_t)}\right) \quad (2)$$

c. Menghitung TF-IDF

$$TF - IDF(t,d) = TF(t,d) \times IDF_t \quad (3)$$

3.12. Splitting Data

Splitting data dibagi menjadi dua bagian, yaitu data untuk pelatihan (training) dan data untuk pengujian (testing). Proses ini dilakukan dengan beberapa variasi rasio pembagian data, dengan data latih dan data uji 80:20 dan 70:30. Pembagian ini dimaksudkan untuk mencari konfigurasi terbaik guna menghasilkan akurasi model yang optimal.[20]

3.13. Imbalance Data

Dataset memiliki distribusi jumlah data antar kelas yang tidak seimbang, kondisi ini dikenal sebagai *imbalance dataset*. Salah satu teknik yang dapat digunakan untuk menangani permasalahan tersebut adalah metode *Random Oversampling*. Metode oversampling bekerja dengan menambah jumlah data pada kelas minoritas melalui proses augmentasi, sehingga proporsi antar kelas menjadi lebih seimbang.[9]

3.14. Modeling

Model klasifikasi yang dibangun menggunakan algoritma SVM dan *Decision Tree* dilatih dengan data training. Setelah proses ini selesai, model diuji menggunakan data testing untuk mengukur kemampuannya dalam mengidentifikasi sentimen dari ulasan yang belum dikenali sebelumnya. Hasil pengujian akan menunjukkan efektivitas masing-masing algoritma.

3.15. Evaluasi

Evaluasi dilakukan dengan menggunakan confusion matrix precision untuk memperlihatkan jumlah prediksi yang benar dan salah. Recall, dan f1-score untuk mengukur ketepatan kemampuan menilai sejauh mana kemampuan model dalam mendeteksi dan membedakan sentimen positif maupun negatif secara lebih mendalam dan akurasi untuk menentukan proporsi prediksi yang benar terhadap keseluruhan jumlah data uji.[6]

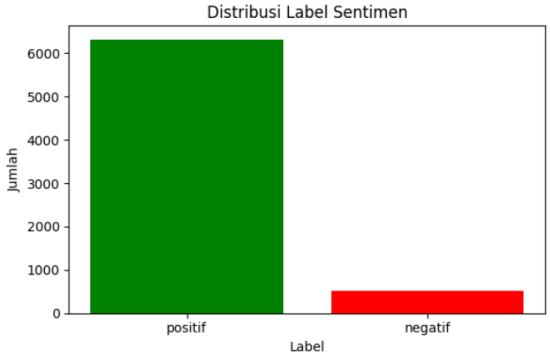
4. HASIL DAN PEMBAHASAN

4.1. Hasil Preprocessing

Tahapan preprocessing diawali dengan pembersihan data untuk memastikan kualitas dataset yang digunakan. Dari total 10.000 data awal yang telah dikumpulkan, dilakukan proses penyaringan dengan mengambil hanya ulasan yang berbahasa Indonesia. Hasilnya, jumlah data berkurang menjadi 6.839 ulasan berbahasa indonesia. Setelah itu, dilakukan penghapusan ulasan ganda, data kosong, teks yang hanya mengandung simbol atau karakter tidak relevan, serta entri dengan struktur teks yang tidak dapat diproses. Selanjutnya, data yang telah bersih melalui proses normalisasi teks, seperti case folding, tokenizing, dan penghapusan kata yang tidak memiliki makna penting. Dataset hasil preprocessing ini kemudian digunakan pada tahap pelabelan sentimen. Contoh hasil pengolahan data dan pelabelan sentimen ditampilkan pada Tabel 2.

Tabel 2. Dataset Seletah Preprocessing dan Labeling

Komentar awal	Komentar akhir	rating	label
Ga akurat	tidak akurat	1	Negatif
Sekarang bnyak iklannya yah duolingo	sekarang banyak iklan duolingo	2	Negatif
Cukup membosankan	cukup bosan	3	Positif
bagus dan bermanfaat	bagus manfaat	4	Positif
sangat membantu sekali	sangat bantu sekali	5	Positif



Gambar 2. Distribusi Label Sentimen

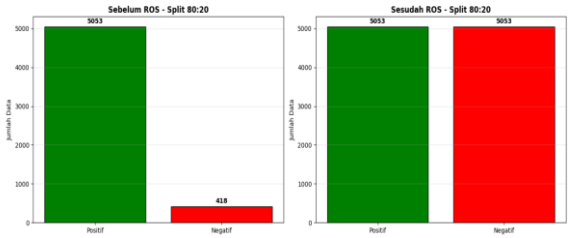
Pada Gambar 2. Menunjukan bahwa adanya ketidakseimbangan data pada kelas positif dan negatif. Data sentimen positif menunjukkan lebih banyak dibandingkan dengan data sentimen negatif dengan masing-masing sebanyak 6.316 data positif dan 523 data negatif. Dari kondisi ini maka di perlukan penyeimbangan data.

4.1. Penyeimbangan Data

Pada Tabel 3 hasil proses penyeimbangan data menggunakan metode *Random Oversampling (ROS)* pada dua skenario rasio pembagian data, yaitu 80:20 dan 70:30. Sebelum dilakukan penyeimbangan, distribusi kelas sentimen menunjukkan kondisi *imbalance*, di mana jumlah data sentimen positif jauh lebih besar dibandingkan sentimen negatif. Setelah diterapkan ROS pada data latih, jumlah data pada masing-masing kelas menjadi seimbang, sementara data uji tetap menggunakan distribusi data asli tanpa penyeimbangan.

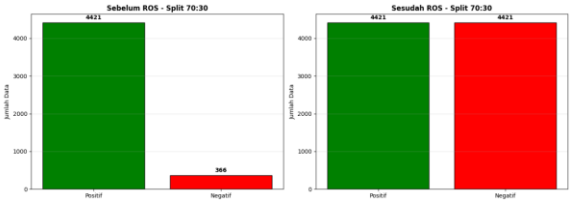
Tabel 3. Distribusi dan Penyeimbangan Data Train dan Data Test

Rasio Split	Kelas sentimen	Jumlah data awal (imbalance)	Jumlah data ros (balance)	Data test (asli)
	Positif	5.053	5.053	1.263
80:20	Negatif	418	5.053	105
	Total Data Train	5.471	10.106	1.368
	Positif	4.421	4.421	1.895
70:30	Negatif	366	4.421	157
	Total Data Train	4.786	8.842	2.052



Gambar 3. Perbandingan Rasio 80:20

Gambar 3 menunjukkan perbandingan distribusi data sentimen pada rasio pembagian 80:20 antara data latih dan data uji. Terlihat bahwa sebelum penyeimbangan, data latih mengalami ketidakseimbangan kelas, khususnya pada sentimen negatif. Setelah dilakukan penyeimbangan menggunakan *Random Oversampling*, jumlah data pada masing-masing kelas sentimen menjadi seimbang, sedangkan data uji tetap mempertahankan distribusi data asli.



Gambar 4. Perbandingan Rasio 70:30

Gambar 4 menggambarkan kondisi pembagian data dengan rasio 70:30, di mana sebagian besar data digunakan sebagai data latih dan sisanya sebagai data uji. Pada tahap awal, terlihat perbedaan jumlah data antar kelas sentimen pada data latih. Setelah dilakukan proses penyeimbangan, proporsi data antar kelas menjadi setara, sehingga data latih lebih representatif untuk digunakan dalam proses pelatihan model, sementara data uji tetap dipertahankan sesuai kondisi aslinya.

4.2. Hasil Pengujian Algoritma

Hasil pengujian algoritma dilakukan untuk mengevaluasi kinerja *Support Vector Machine* (SVM) dan *Decision Tree* dalam mengklasifikasikan sentimen ulasan pengguna *Duolingo*. Pengujian dilakukan menggunakan dua skenario rasio pembagian data, yaitu 80:20 dan 70:30, dengan kondisi data latih *imbalance* dan *balance* menggunakan *Random Oversampling* (ROS). Tujuannya adalah untuk membandingkan kinerja SVM dan *Decision Tree*.

4.3. Hasil Pengujian Rasio Split 80:20

Pada pengujian ini, sebanyak 80% data digunakan sebagai data latih dan 20% sisanya sebagai data uji. Hasil perbandingan performa model antara kondisi data latih yang tidak seimbang (tanpa ROS) dan data latih yang telah diseimbangkan (menggunakan ROS) pada skenario rasio 80:20 disajikan pada Tabel 4.

Tabel 4. Hasil Komparasi Rasio Split 80:20

Algoritma	Kondisi Train	Akurasi	Pecision	Recall	F1-Score
SVM	Imbalance	93.57%	92.52%	93.57%	92.11%
SVM	Balance	92.76%	92.99%	92.76%	92.87%
Decision Tree	Imbalance	91.30%	90.17%	91.30%	90.65%
Decision Tree	Balance	90.35%	90.85%	90.35%	90.59%

4.4. Hasil Pengujian Rasio Split 70:30

Pada pengujian ini, sebanyak 70% data digunakan sebagai data latih dan 30% sisanya sebagai data uji. Perbandingan kinerja model pada kondisi data latih yang tidak seimbang (tanpa ROS) dan data latih yang telah diseimbangkan (menggunakan ROS) pada skenario rasio 70:30 disajikan pada Tabel 5.

Tabel 5. Hasil Komparasi Rasio Split 70:30

Algoritma	Kondisi Train	Akurasi	Pecision	Recall	F1-Score
SVM	Imbalance	93.13%	91.69%	93.13%	91.72%
SVM	Balance	92.74%	92.97%	92.74%	92.85%
Decision Tree	Imbalance	91.86%	91.36%	91.86%	91.59%
Decision Tree	Balance	90.25%	91.30%	90.25%	90.73%

4.5. Model Optimal dan Analisis Rasio Split

Mengacu pada hasil pengujian yang tercantum pada Tabel 4 dan Tabel 5, algoritma *Support Vector Machine* (SVM) dengan rasio pembagian data 80:20 pada kondisi *imbalance* menunjukkan performa terbaik berdasarkan nilai akurasi tertinggi sebesar 93,57% serta F1-score sebesar 92,11%. Namun demikian, kemampuan model dalam mengidentifikasi ulasan negatif masih belum optimal, yang tercermin dari rendahnya nilai *recall* pada kelas negatif. Kondisi ini menyebabkan sebagian ulasan yang mengandung sentimen negatif belum berhasil terdeteksi secara memadai. Mengingat bahwa ulasan negatif memiliki peranan penting bagi pengembang dalam mengidentifikasi permasalahan dan meningkatkan kualitas layanan aplikasi, maka diperlukan model yang mampu mengenali kelas tersebut dengan lebih baik. Penerapan metode *Random Oversampling* (ROS) memberikan kontribusi dalam meningkatkan performa pada kelas minoritas, khususnya dalam memperbaiki nilai *recall*. Oleh karena itu, penentuan model terbaik tidak hanya mempertimbangkan nilai akurasi, tetapi juga memperhatikan keseimbangan kemampuan model dalam mengklasifikasikan setiap kelas sentimen sesuai dengan tujuan penelitian.

Penentuan model optimal dalam penelitian ini tidak hanya didasarkan pada pencapaian akurasi tertinggi, tetapi juga pada kemampuan model dalam mendeteksi kelas minoritas, khususnya sentimen negatif yang memiliki nilai praktis penting bagi pengembang aplikasi. Hasil pengujian menunjukkan bahwa SVM pada rasio 80:20 tanpa penyeimbangan data menghasilkan akurasi tertinggi sebesar 93,57%, namun masih memiliki keterbatasan dalam mengenali sentimen negatif yang tercermin dari rendahnya nilai recall. Penerapan *Random Oversampling* (ROS) pada data latih menyebabkan penurunan akurasi global, tetapi secara signifikan meningkatkan recall dan F1-score pada kelas negatif, sehingga menghasilkan model yang lebih seimbang. Oleh karena itu, SVM dengan penerapan ROS dinilai lebih relevan secara fungsional, dengan pemilihan model terbaik mempertimbangkan trade-off antara akurasi dan sensitivitas terhadap kelas minoritas, bukan semata-mata akurasi global.

4.6. Dampak Ros dan Trade-off Kinerja

Penerapan metode Random Oversampling (ROS) pada data latih menyebabkan penurunan nilai akurasi dan F1-score pada seluruh skenario pengujian, sehingga model dengan kondisi data *imbalance* masih unggul dari sisi akurasi keseluruhan. Sebagai contoh, pada algoritma *Support Vector Machine* (SVM) dengan rasio split 80:20, nilai akurasi menurun dari 93,57% menjadi 92,76% setelah penerapan ROS. Meskipun demikian, ROS berhasil mengurangi bias model terhadap kelas mayoritas dengan meningkatkan kemampuan klasifikasi pada kelas minoritas. Kondisi ini menunjukkan adanya *trade-off* kinerja, di mana penurunan akurasi global dikompensasikan dengan peningkatan sensitivitas model terhadap kelas negatif. *Trade-off* tersebut merupakan konsekuensi yang wajar dalam penanganan data tidak seimbang, karena menghasilkan model yang lebih proporsional dan lebih sesuai dengan tujuan analisis sentimen.

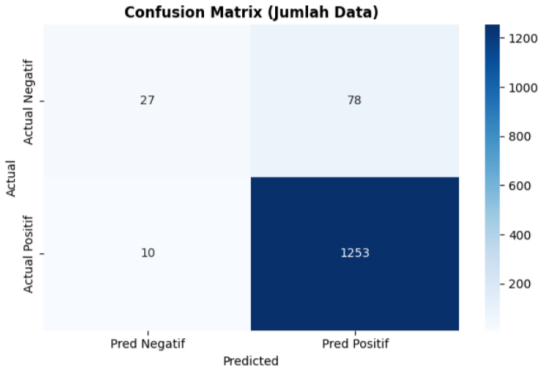
Tabel 6. Komparasi Kinerja Kelas Negatif

Algoritma	Split Rasio	Kondisi Train	Recall	F1-Score
SVM	80:20	Imbalance	25.71%	38.03%
		Balance	56.19%	54.38%
	70:30	Imbalance	24.84%	35.62%
		Balance	56.05%	54.15%
Decision Tree	80:20	Imbalance	29.52%	34.25%
		Balance	42.86%	40.54%
	70:30	Imbalance	40.13%	43.00%
		Balance	48.41%	43.18%

4.7. Interpretasi Matrix dan Confusion Matrix

Pembahasan ini mengacu pada Gambar 5 yang menyajikan *Confusion Matrix* dari model *Support Vector Machine* (SVM) pada rasio pembagian data 80:20 dalam kondisi *imbalance*, di mana model tersebut menghasilkan nilai akurasi tertinggi sebesar 93,57% yang terutama dipengaruhi oleh tingginya jumlah *True Positive (TP)* sehingga sebagian besar ulasan positif dapat diklasifikasikan dengan benar. Namun demikian, hasil *Confusion Matrix* juga menunjukkan keterbatasan model yang tercermin dari tingginya nilai *False Positive (FP)*, yang mengindikasikan masih adanya ulasan negatif yang keliru diklasifikasikan sebagai positif, sehingga berdampak pada rendahnya nilai *recall* pada kelas negatif. Mengingat sentimen negatif merupakan kelas minoritas yang memiliki nilai praktis penting bagi pengembang aplikasi, akurasi global tidak dijadikan satu-satunya dasar dalam penentuan model terbaik. Oleh karena itu, metode *Random Oversampling* (ROS) diterapkan untuk menyeimbangkan distribusi data latih, yang meskipun menyebabkan penurunan akurasi secara keseluruhan, terbukti mampu meningkatkan nilai *recall* dan *F1-score* pada kelas negatif secara signifikan. Dengan demikian, SVM dengan penerapan ROS dinilai sebagai model yang lebih optimal dan

relevan secara fungsional karena mampu mencapai keseimbangan antara akurasi dan sensitivitas terhadap kelas minoritas, bukan semata-mata berfokus pada pencapaian akurasi tertinggi.



Gambar 5. Confusion Matrix Mode Optimal

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Support Vector Machine* (SVM) menunjukkan performa yang lebih unggul dibandingkan *Decision Tree* pada seluruh skenario pengujian. Pada kondisi data tidak seimbang dengan pembagian data 80:20, SVM menghasilkan akurasi tertinggi sebesar 93,57%, namun masih memiliki keterbatasan dalam mendeteksi kelas minoritas, yang ditunjukkan oleh nilai recall sentimen negatif yang rendah sebesar 25,71%. Setelah dilakukan penyeimbangan data menggunakan *Random Oversampling* (ROS), meskipun akurasi global SVM sedikit menurun menjadi 92,76%, terjadi peningkatan yang signifikan pada kemampuan model dalam mengenali sentimen negatif dengan recall sebesar 56,19% dan F1-score sebesar 54,38%, sehingga menghasilkan model yang lebih seimbang dan sensitif terhadap kelas minoritas dibandingkan *Decision Tree*, yang hanya mencapai akurasi tertinggi sebesar 91,86% dan F1-score kelas negatif sebesar 43,18%. Berdasarkan hasil analisis sentimen, komentar positif pengguna *Duolingo* didominasi oleh apresiasi terhadap kemudahan dan efektivitas aplikasi dalam membantu proses pembelajaran bahasa, khususnya bahasa Inggris, sedangkan komentar negatif sebagian besar menyoroti iklan yang berlebihan serta penurunan kualitas aplikasi dibandingkan versi sebelumnya. Dengan mempertimbangkan bahwa penentuan model optimal tidak hanya didasarkan pada pencapaian akurasi tertinggi, tetapi juga pada kemampuan model dalam mendeteksi sentimen negatif yang memiliki nilai praktis penting, maka SVM dengan penerapan *Random Oversampling* pada pembagian data 80:20 dinilai sebagai model yang paling relevan secara fungsional dalam penelitian ini. Untuk pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi metode penyeimbangan data lain seperti SMOTE, menerapkan algoritma klasifikasi tambahan seperti *Random Forest* atau pendekatan Deep Learning, menambahkan kategori sentimen netral, serta memperluas dan memperkaya dataset ulasan guna meningkatkan kemampuan generalisasi model dalam memprediksi persepsi pengguna.

DAFTAR PUSTAKA

[1] I. Duolingo, “Duolingo surpasses 50 million daily active users and grows revenue in 2025,” Duolingo Investor Relations. [Online]. Available: https://investors.duolingo.com

[2] Y. Finance, “Duolingo surpasses 50 million daily active users,” Yahoo Finance. [Online]. Available: https://finance.yahoo.com/news/duolingo-surpasses-50-million-daily-210100511.html

[3] J. Empati, T. Salsabila, N. Nafilah, F. Patangga, S. Zulfa, and N. Listyaningsih, “LITERATURE REVIEW : EFEKTIVITAS PENGGUNAAN APLIKASI DUOLINGO TERHADAP MOTIVASI BELAJAR,” vol. 13, pp. 302–312, 2024.

[4] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, “Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine,” *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.

[5] S. Putri, A. Ati, P. Muharani, and R. Qori, “Sentiment Analysis of Gojek User Reviews using TF-IDF and Machine Learning in Orange Platform,” vol. 6, no. 4, pp. 296–305, 2025.

[6] E. F. Laili *et al.*, “KOMPARASI ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE (SVM) DALAM,” vol. 8, no. 1, pp. 67–76, 2025.

[7] F. Alifiana, M. F. Asnawi, I. A. Ihsannudin, M. A. M. Baihaqy, and D. Asmarajati, “Analisis Sentimen Aplikasi Duolingo Menggunakan Algoritma Naïve Bayes Dan Support Machine Learning,” *Device*, vol. 13, no. 2, pp. 223–230, 2023, doi: 10.32699/device.v13i2.5905.

[8] K. A. Rokhman, B. Berlilana, and P. Arsi, “Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online,” *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021, doi: 10.24076/joism.2021v3i1.341.

[9] Z. Abidin, T. Suratno, and M. F. Putri, “PENERAPAN RANDOM OVERSAMPLING DAN PRINCIPAL COMPONENT ANALYSIS UNTUK MENINGKATKAN AKURASI PREDIKSI KEBANGKRUTAN APPLICATION OF RANDOM OVERSAMPLING AND PRINCIPAL COMPONENT ANALYSIS TO ENHANCE THE ACCURACY OF BANKRUPTCY PREDICTION FOR,” vol. 12, no. 5, 2025.

[10] T. A. Anastasya, A. Diani, P. Saka, M. Juventus, D. Deke, and A. M. Rizki, “OPTIMASI ALGORITMA SVM DENGAN PSO UNTUK

ANALISIS SENTIMEN PADA MEDIA SOSIAL X TERHADAP KINERJA TIMNAS DI ERA SHIN TAE-YONG," vol. 9, no. 1, pp. 384–391, 2025.

[11] 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.

[12] U. Indahyanti *et al.*, "Perbandingan Algoritma Machine Learning dalam," vol. 11, no. 2, pp. 96–105, 2025.

[13] Z. Y. Burnama, M. A. Rosid, and N. L. Azizah, "Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning Sentiment Analysis on YouTube Comments in MPL Season 13 Tournament Using Ensemble Machine Learning Method".

[14] A. Zhahrina, U. Sofiah, D. Wahyu, A. Andayani, and N. Khairurrabbani, "Penerapan Metode Support Vector Machine (SVM) dalam Analisis Sentimen Ulasan Aplikasi Tokopedia di Google Play Store," no. September 2025, pp. 84–91.

[15] A. Okta, K. Adi, A. Sanjaya, and A. B. Setiawan, "Penerapan Inset Lexicon untuk Analisis Sentimen Penonton Video JKT48 di YouTube," vol. 9, pp. 1276–1283.

[16] J. Penerapan, T. Informasi, D. S. Sarassati, S. Yulianto, J. Prasetyo, and B. Rob, "IT-EXPLORE Analisis sentimen terhadap dampak banjir rob dengan menggunakan metode Support Vector Machine," vol. 04, pp. 233–244, 2025, doi: 10.24246/itexplore.v4i2.2025.pp233-244.

[17] R. A. P. Sari, S. Kacung, and B. Santoso, "ANALISIS SENTIMEN LAYANAN KESEHATAN BPJS MENGGUNAKAN METODE SVM," pp. 878–885, 2025.

[18] D. Fitriono, R. Indriati, and A. Ristyawan, "Analisis Sentimen Ulasan Aplikasi Chatgpt Menggunakan Algoritma Support Vector Machine dan Lexicon Based," vol. 3, no. 2, pp. 101–111, 2025.

[19] S. V. M. Tf-idf, "Analisis Sentimen terhadap RSUD Salatiga Menggunakan Abstrak," vol. 6, no. 1, pp. 478–489, 2025.

[20] F. M. Fadillah, Y. Cahyana, and A. Fauzi, "BULLETIN OF COMPUTER SCIENCE RESEARCH Analisis Sentimen Masyarakat Terhadap Pembatasan BBM Pertalite Menggunakan Random Forest dan K-Nearest Neighbor," vol. 5, no. 4, pp. 340–352, 2025, doi: 10.47065/bulletincsr.v5i4.547.