

Analisis Komparatif Metode Logistic Regression dan Support Vector Machine pada Sentimen Aplikasi GuitarTuna

Comparative Analysis of Logistic Regression and Support Vector Machine Methods for GuitarTuna App Sentiment

Muhammad Yusuf, M Rudi Sanjaya*, Dwi Rosa Indah, Ardina Ariani

Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Sriwijaya, Indralaya, Indonesia

Email: ¹muhammadyusuf25632@gmail.com, ^{2,*}m.rudi.sjy@ilkom.unsri.ac.id, ³indah812@gmail.com, ⁴ardinaariani@unsri.ac.id

Email Penulis Korespondensi: m.rudi.sjy@ilkom.unsri.ac.id

Received: 09/09/2026, Revised: 24/09/2026, Accepted: 27/09/2026, Published: 30/09/2026

Abstrak—Studi ini difokuskan untuk mengidentifikasi kecenderungan sentimen pengguna aplikasi GuitarTuna, sekaligus melakukan komparasi performa antara algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Sebanyak 20.339 data mentah berhasil dihimpun melalui Google Play Store, App Store, serta kuesioner. Setelah melalui proses penyaringan yang ketat, meliputi eliminasi data ganda, teks kosong, dan respons tidak relevan, diperoleh 13.615 data valid. Rangkaian analisis kemudian dilanjutkan dengan tahap *text-preprocessing*, ekstraksi fitur berbasis pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF), hingga proses klasifikasi menggunakan kedua algoritma tersebut. Hasil evaluasi menunjukkan bahwa SVM mencapai *accuracy* sebesar 90,53%, *macro recall* 0,60, dan *macro F1-score* 0,62. Performa ini sedikit lebih unggul secara deskriptif dibandingkan LR yang mencatatkan *accuracy* 90,30%, *macro recall* 0,56, dan *macro F1-score* 0,59. Walaupun demikian, pengujian *McNemar* yang menghasilkan *p-value* sebesar 0,6338 membuktikan bahwa keunggulan nominal SVM tersebut tidak memiliki signifikansi secara statistik. Lebih lanjut, keterbatasan jumlah sampel pada kelas netral turut berdampak pada rendahnya tingkat presisi prediksi kedua model pada kategori tersebut. Dari aspek topik ulasan, sentimen positif sangat didorong oleh kepuasan pengguna terhadap presisi *tuning*, kemudahan operasional, dan fitur edukasi. Sebaliknya, sentimen negatif didominasi oleh keluhan terkait biaya langganan, hambatan transaksi, pembatasan akses fitur, dan masalah konektivitas. Adapun kontribusi utama dari penelitian ini mencakup dua aspek. Secara praktis, temuan dari ekstraksi topik ini menyajikan panduan evaluasi yang konkret bagi pihak pengembang untuk memprioritaskan perbaikan fitur dan layanan. Sementara secara akademis, studi ini menyumbangkan bukti empiris yang komprehensif terkait evaluasi komparatif algoritma machine learning pada dataset ulasan lintas platform yang berhadapan dengan tantangan ketidakseimbangan kelas.

Kata Kunci: Analisis Sentimen; GuitarTuna; TF-IDF; Logistic Regression; SVM

Abstract—This study focuses on identifying sentiment trends among GuitarTuna app users while comparing the performance of Support Vector Machine (SVM) and Logistic Regression (LR) algorithms. A total of 20,339 raw data points were collected from the Google Play Store, the App Store, and questionnaires. Following a rigorous filtering process, which involved eliminating duplicate entries, empty text, and irrelevant responses, 13,615 valid data points were obtained. The analysis proceeded through text preprocessing, feature extraction using Term Frequency-Inverse Document Frequency (TF-IDF) weighting, and classification using both algorithms. Evaluation results indicate that SVM achieved an accuracy of 90.53%, a macro recall of 0.60, and a macro F1-score of 0.62. This performance was descriptively superior to that of LR, which recorded an accuracy of 90.30%, a macro recall of 0.56, and a macro F1-score of 0.59. However, McNemar's test, yielding a p-value of 0.6338, demonstrated that SVM's nominal advantage was not statistically significant. Furthermore, the limited sample size for the neutral class contributed to low prediction precision for both models within that category. Regarding review topics, positive sentiment was largely driven by user satisfaction with tuning precision, ease of use, and educational features. Conversely, negative sentiment was dominated by complaints concerning subscription costs, transaction issues, feature access restrictions, and connectivity problems. This research makes contributions in two key areas. From a practical standpoint, the topic extraction findings provide developers with concrete guidance for prioritizing feature and service improvements. From an academic perspective, the study offers comprehensive empirical evidence regarding the comparative evaluation of machine learning algorithms on cross-platform review datasets characterized by class imbalance.

Keywords: Sentiment Analysis; GuitarTuna; TF-IDF; Logistic Regression; SVM

1. PENDAHULUAN

Transformasi digital di sektor kreatif, terkhusus pada industri musik, telah mendisrupsi metode pembelajaran instrumen menjadi lebih modern. Fleksibilitas ruang dan waktu kini dapat dicapai secara optimal melalui konsep mobile learning yang mengandalkan perangkat seluler [1]. GuitarTuna hadir sebagai salah satu instrumen edukasi digital terpopuler, yang mengombinasikan fungsionalitas presisi tuning dawai dengan ragam fitur interaktif seperti metronom dan pustaka *chord*. Masifnya jumlah pengguna aplikasi ini memicu lonjakan *user-generated* content dalam bentuk ulasan teks di Google Play Store maupun App Store. Kumpulan opini ini menjadi tambang informasi berharga yang memotret realitas pengalaman pengguna, mulai dari kepuasan terhadap keakuratan penyeteman hingga keluhan teknis mengenai kendala antarmuka, limitasi fitur, dan mahalnnya sistem *subscription* [2]. Kendati demikian, klasifikasi sentimen bervolume raksasa secara manual dihadapkan pada kendala inefisiensi waktu, keterbatasan tenaga, dan rentan terhadap bias pengamat [3]. Guna menanggulangi hambatan tersebut, penelitian ini mengusulkan pendekatan otomatisasi analisis sentimen berbasis *Natural Language Processing* (NLP) menggunakan ekstraksi fitur *Term*

Frequency-Inverse Document Frequency (TF-IDF) yang dieksekusi melalui algoritma *Support Vector Machine* (SVM) dan *Logistic Regression*.

Secara konseptual, analisis sentimen berfungsi untuk membedah opini tak terstruktur menjadi polaritas kelas spesifik, yakni sentimen positif, netral, maupun negatif [4]. Agar korpus teks ulasan dapat dianalisis oleh algoritma *machine learning*, representasi numerik diwujudkan melalui metode TF-IDF. Pendekatan ini memberikan nilai bobot pada sebuah *term* berdasarkan rasio kemunculannya dalam dokumen dan korpus, sehingga kata kunci yang menjadi representasi sentimen akan mendapat atensi lebih tinggi [5]. Untuk proses klasifikasi, *Logistic Regression* dipilih karena efisiensi komputasinya dalam menangani data teks [6]. Di sisi lain, algoritma SVM digunakan karena kemampuannya dalam memisahkan data pada ruang fitur berdimensi tinggi [7]. Komparasi kedua metode komputasi ini diharapkan mampu menyajikan evaluasi empiris yang komprehensif, khususnya dalam menangani ulasan aplikasi musik yang sarat akan istilah teknis dan jargon spesifik.

Implementasi komparasi algoritma analisis sentimen sejatinya telah banyak dieksplorasi lintas domain. Budianto dan kawan-kawan [7] mengevaluasi kemampuan *Logistic Regression* dan SVM pada data ulasan *e-commerce*, di mana SVM menunjukkan akurasi yang lebih superior. Namun, literatur tersebut belum sepenuhnya merepresentasikan konteks penelitian ini karena domain transaksional jual-beli memiliki karakteristik linguistik yang kontras dengan aplikasi musik. Dalam ranah instrumen digital, Ariq [2] secara khusus menyoroti evaluasi aplikasi GuitarTuna dan Fender Guitar Tuner dengan menggunakan kombinasi metode *Naïve Bayes* dan SVM. Penelitian tersebut menunjukkan bahwa ulasan pengguna pada aplikasi penyeter musik dapat digunakan sebagai bahan evaluasi produk, dengan SVM memperoleh performa yang lebih tinggi pada metrik yang dilaporkan. Namun, keterbatasan dari studi ini adalah tidak adanya pengujian banding dengan algoritma regresi linear seperti *Logistic Regression*, serta pengumpulan data yang hanya bergantung pada hasil ekstraksi otomatis satu sumber platform tanpa divalidasi oleh data primer seperti kuesioner pengguna, sehingga representasi sentimen yang dihasilkan berpotensi kurang holistik.

Pada literatur berbahasa Indonesia, Lestari dan Hutagalung [5] menguji efektivitas metode pembobotan TF-IDF untuk analisis sentimen menggunakan *Support Vector Machine*, *Logistic Regression*, dan *Naïve Bayes*. Fokus kajian mereka diarahkan pada optimasi ekstraksi fitur teks pada ulasan aplikasi jasa pengiriman. Hasil pengujian menyimpulkan bahwa implementasi TF-IDF secara signifikan mampu meningkatkan kemampuan deteksi klasifikasi algoritma *machine learning*. Akan tetapi, batasan dalam riset tersebut terletak pada domain *dataset* logistik yang secara linguistik memiliki perbandingan kelas kata yang jauh berbeda dengan ulasan pada aplikasi mobile learning sektor musik, di mana pengguna GuitarTuna kerap kali menyertakan istilah teknis seperti *chord*, dawai, dan frekuensi nada.

Penelitian lain oleh Maulana et al. [8] menerapkan SVM untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi finansial Pluang guna memantau indeks kepuasan investor secara real-time. Penelitian ini berfokus pada pemetaan sentimen negatif sebagai upaya mitigasi risiko layanan pelanggan. Studi ini menunjukkan bahwa ulasan publik digital dapat digunakan sebagai sumber informasi mengenai pengalaman pengguna. Namun, keterbatasan utama dari studi finansial ini adalah minimnya penanganan penyeimbangan data pada kelas sentimen netral, serta absennya benchmarking dengan model klasifikasi lain, sehingga tidak dapat dipastikan apakah SVM adalah model paling efisien untuk *dataset* berskala besar tersebut.

Sebagai tambahan referensi, Cahyani dan Prasetyaningrum [9] mengeksplorasi spektrum algoritma dengan membandingkan SVM, *Logistic Regression*, dan *Random Forest* dalam menganalisis sentimen pada ulasan aplikasi berbasis kecerdasan buatan. Kajian ini menitikberatkan pada perolehan akurasi tertinggi antara pendekatan *ensemble learning* dibandingkan algoritma linear. Temuannya mengindikasikan bahwa perbandingan banyak algoritma sangat penting untuk menemukan kecocokan (*best fit*) terhadap karakteristik *dataset*. Namun, penelitian tersebut belum menyajikan penjabaran *confusion matrix* secara detail untuk mengevaluasi bias distribusi kelas, dan sama sekali tidak menyentuh tantangan *preprocessing* pada data aplikasi instrumen musik yang sering mengandung terminologi dwibahasa (Inggris-Indonesia).

Berdasarkan tinjauan terhadap kelima penelitian tersebut, masih terdapat ruang penelitian (*research gap*) yang berkaitan dengan komparasi SVM dan *Logistic Regression* pada ulasan aplikasi GuitarTuna. Untuk mengisi kesenjangan data dari penelitian sebelumnya, riset ini mengintegrasikan data sekunder (ulasan App Store dan Google Play Store) dengan data primer yang dijamin melalui instrumen kuesioner. Populasi kuesioner ditetapkan pada siswa dan alumni SMA Negeri 1 Indralaya. Penarikan sampel kemudian dikerucutkan menggunakan teknik *purposive sampling*, yang secara spesifik menyoroti anggota dan alumni Komunitas Seni Arta Bhavana dengan kriteria inklusi berupa rekam jejak interaksi langsung terhadap aplikasi GuitarTuna. Interaksi ini mencakup penggunaan melalui perangkat pribadi maupun penggunaan bersama saat sesi latihan komunitas. Mengingat kuesioner ini dirancang untuk mengumpulkan korpus teks opini terbuka (sebagai *user-generated content*) alih-alih mengukur skala psikometrik, validitas instrumen didekati melalui *face validity* (validitas rupa) melalui evaluasi dan persetujuan dosen pembimbing selaku pakar. Keputusan pakar memastikan bahwa instruksi terbuka pada kuesioner telah relevan untuk memancing responden mengartikulasikan pengalamannya terhadap fitur aplikasi. Lebih lanjut, kualitas data dikendalikan melalui mekanisme *data cleaning*, di mana respons tidak relevan, kosong, maupun duplikat secara sistematis dieliminasi sebelum tahapan ekstraksi fitur dilakukan. Proses tersebut ditujukan untuk meningkatkan konsistensi data yang digunakan dalam analisis, tetapi tidak sepenuhnya menghilangkan kemungkinan bias seleksi akibat pengurangan data.

Secara keseluruhan, studi ini bertujuan mengklasifikasikan polaritas sentimen pengguna aplikasi GuitarTuna sekaligus menyajikan pembuktian empiris terkait perbandingan performa *Logistic Regression* dan SVM berbasis TF-

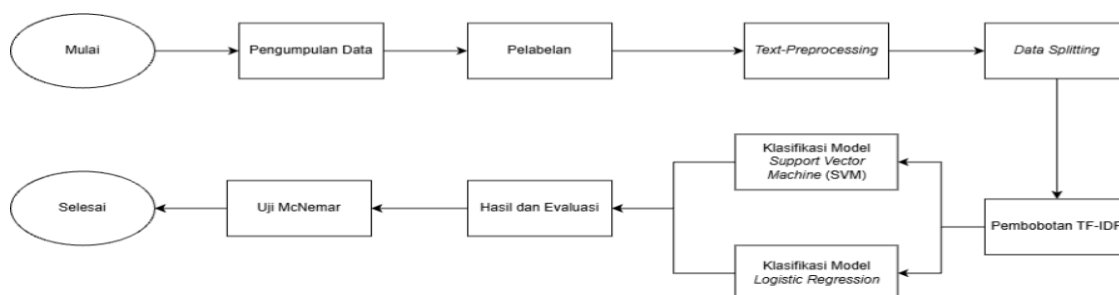
IDF. Evaluasi performa model diukur menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk memetakan konsistensi algoritma pada tiap kelas polaritas. Integrasi sumber data yang hibrida, serta metodologi *purposive sampling* yang digunakan sesuai dengan kriteria inklusi penelitian, diharapkan tidak hanya mengurai preferensi sentimen pengguna GuitarTuna, tetapi juga memberikan referensi tambahan mengenai klasifikasi teks pada domain aplikasi pembelajaran instrumen digital.

Secara keseluruhan, studi ini bertujuan mengklasifikasikan polaritas sentimen pengguna aplikasi GuitarTuna sekaligus menyajikan pembuktian empiris terkait perbandingan performa algoritma *Logistic Regression* dan SVM berbasis TF-IDF. Evaluasi model diukur menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* guna memetakan konsistensi algoritma pada tiap kelas polaritas. Integrasi sumber data hibrida dan metodologi *purposive sampling* yang diterapkan sesuai kriteria inklusi diharapkan mampu mengurai preferensi pengguna secara lebih holistik. Sebagai pelengkap dari tujuan tersebut, penelitian ini menawarkan kontribusi pada dua ranah utama. Secara praktis, hasil ekstraksi topik dan pemetaan sentimen ini menyajikan panduan evaluasi yang konkret bagi pihak pengembang aplikasi untuk memprioritaskan perbaikan fitur, penyesuaian biaya berlangganan, dan peningkatan kualitas layanan berdasarkan masukan pengguna. Secara akademis, studi ini menyumbangkan bukti empiris yang komprehensif mengenai evaluasi komparatif algoritma *machine learning* pada dataset ulasan lintas platform yang dihadapkan pada tantangan ketidakseimbangan kelas (*imbalanced data*), sekaligus memperkaya literatur pemrosesan bahasa alami (*Natural Language Processing*) pada domain aplikasi instrumen musik digital.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Riset ini mengadopsi kerangka kerja kuantitatif dengan alur terstruktur, bermula dari ekstraksi data mentah hingga pengujian efektivitas algoritma. Seluruh proses komputasi dilakukan menggunakan Google Colab dan bahasa pemrograman *Python*. Eksperimen ini didesain guna mengevaluasi serta membandingkan secara komprehensif reliabilitas *Support Vector Machine* (SVM) dan *Logistic Regression* dalam menganalisis sentimen pengguna GuitarTuna. Untuk menjaga objektivitas komparasi, kedua model diuji pada korpus data dan perlakuan pra-pemrosesan yang sepenuhnya identik. Visualisasi alur kerangka kerja tersebut dijabarkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan diagram yang disajikan pada gambar 1, tahapan awal penelitian ini adalah menghimpun ulasan pengguna dari Google Play Store, App Store, serta kuesioner. Dari total 20.339 data mentah yang diperoleh, dilakukan proses pembersihan untuk mengeliminasi duplikasi dan komentar tanpa isi, sehingga menyisakan 13.615 data bersih. Langkah berikutnya mencakup pemberian label sentimen yang mengacu pada nilai *rating*, diikuti dengan prapemrosesan teks (*text-preprocessing*). *Dataset* yang telah siap ini kemudian dipartisi menjadi 80% data latih dan 20% data uji melalui teknik *stratified split*. Setelah dipisahkan, ekstraksi fitur TF-IDF diimplementasikan pada data latih, dan model pembobotan tersebut diaplikasikan untuk mengubah format data uji. Vektor TF-IDF ini selanjutnya bertindak sebagai data masukan bagi algoritma klasifikasi *Logistic Regression* dan *Support Vector Machine* (SVM). Kinerja dari kedua model tersebut diukur menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, beserta *confusion matrix*. Terakhir, pengujian statistik McNemar diterapkan guna mengevaluasi signifikansi perbedaan hasil prediksi antara kedua model klasifikasi tersebut.

2.2 Objek Penelitian

Fokus kajian ini bertumpu pada korpus teks ulasan konsumen aplikasi GuitarTuna. Pemilihan instrumen lunak ini dilatarbelakangi oleh kapabilitasnya sebagai utilitas *tuning* dawai terpadu yang juga menyediakan modul edukasi, sehingga memancing ragam spektrum sentimen terkait kualitas antarmuka maupun layanannya.

Eksplorasi data disokong oleh tiga sumber utama. Selain penambahan sekunder dari platform distribusi iOS dan Android, studi ini juga mengakuisisi data primer yang menyasar subjek siswa dan alumni SMA Negeri 1 Indralaya. Penjaringan sampel dikhususkan pada anggota serta alumni Komunitas Seni Arta Bhavana melalui teknik *purposive sampling*. Kriteria inklusinya mewajibkan responden memiliki riwayat interaksi empiris dengan GuitarTuna, tanpa membatasi kepemilikan gawai, artinya, pengalaman pemakaian secara komunal saat agenda

komunitas tetap diakui valid. Pengumpulan data primer ini dilakukan menggunakan instrumen kuesioner untuk menjangkau *rating* sekaligus deskripsi pengalaman tekstual yang mendalam.

2.3 Pengumpulan Data

Tujuan utama pada fase ini adalah memetakan persepsi konsumen saat mengeksplorasi fitur penyetem maupun sarana belajar GitarTuna secara mandiri. Entri opini yang diserap merangkum apresiasi pengguna (seperti kepraktisan dan akurasi frekuensi) hingga berbagai keluhan teknis (kendala pembayaran *subscription* hingga limitasi fitur).

Guna memfasilitasi komputasi, korpus diorganisasikan secara sistematis ke dalam format tabular (Tabel 1). Variabel *rating* difungsikan sebagai referensi pembentukan kelas polaritas, sedangkan *review_text* diplot sebagai komponen utama bagi ekstraksi fitur NLP. Di sisi lain, atribut pendukung dijaga guna menjamin *traceability* sumber *dataset*.

Tabel 1. Deskripsi Atribut *Dataset* Ulasan GitarTuna

Nama Kolom	Tipe Data	Deskripsi Atribut
<code>review_id</code>	integer	ID unik dari setiap ulasan pengguna.
<code>date_created</code>	object	Tanggal dan waktu ulasan dipublikasikan.
<code>rating</code>	integer	Tingkat <i>rating</i> bintang (1 hingga 5) dari pengguna.
<code>review_text</code>	object	Teks lengkap ulasan terkait pengalaman menggunakan aplikasi.
<code>source_platform</code>	object	Sumber asal data (Google Play Store, App Store, atau Kuesioner)
<code>version</code>	object	Versi aplikasi GitarTuna saat ulasan diberikan.

Tabel 1 menunjukkan atribut yang digunakan dalam *dataset* penelitian. Dari seluruh atribut tersebut, *rating* dan *review_text* menjadi atribut utama dalam proses klasifikasi sentimen, sedangkan atribut lainnya digunakan sebagai informasi pendukung dan identifikasi sumber data. Data yang telah memenuhi kriteria kemudian digunakan sebagai masukan untuk tahap *preprocessing* dan pemodelan.

Agregasi data sekunder dikurasi dari App Store dan Google Play berbekal layanan AppFollow, sementara himpunan data primer dijangkau melalui platform Google Form yang disalurkan kepada populasi target kriteria [10]. Konsolidasi seluruh sumber data menghasilkan 20.339 entri awal. Selanjutnya, dilakukan proses *data cleaning* untuk mengeliminasi entri duplikat, teks kosong, dan respons yang tidak relevan dengan objek penelitian. Proses tersebut menghasilkan 13.615 data bersih yang digunakan pada tahap pelabelan, prapemrosesan, pembobotan TF-IDF, dan pemodelan klasifikasi. Pengurangan data dilakukan berdasarkan kriteria kualitas dan relevansi, bukan berdasarkan label sentimen, meskipun kemungkinan bias seleksi akibat proses penyaringan tetap menjadi keterbatasan penelitian. Perubahan ukuran data tersebut berdasar pada kriteria kualitas dan relevansi yang mengakibatkan beberapa data itu dibersihkan, pengurangan data tersebut dapat diamati pada tabel berikut.

Tabel 2. Perubahan Ukuran Data

Proses	Jumlah Data
Data Awal	20.339
Data Dihapus karena Duplikat	4.300
Data Dihapus karena Kosong / Tidak Relevan	2.424
Jumlah Data Akhir	13.615

Tabel 2 menjelaskan bahwa penurunan jumlah data dari 20.339 entri awal menjadi 13.615 data bersih menunjukkan bahwa proses *cleaning* menghapus data duplikat, data kosong, dan data yang tidak relevan. Jumlah data tersebut berkurang sekitar 33% dari keseluruhan data awal. Pada *dataset* penelitian ini, penghapusan data duplikat dan data kosong dilakukan untuk mengurangi data yang tidak diperlukan sebelum proses ekstraksi fitur TF-IDF. Dengan demikian, data yang digunakan pada tahap pemodelan merupakan data yang telah melalui proses penyaringan berdasarkan kriteria yang telah ditetapkan.

2.4 Pre-Processing

Rangkaian prapemrosesan diaplikasikan untuk merekonstruksi karakter mentah menjadi struktur informasi yang terstruktur dan analitis [11]. Standarisasi ini dibuka dengan *case folding*, yang menyeragamkan penggunaan huruf dengan mengubah seluruh leksikon ke dalam format huruf kecil (*lowercase*). Tahap *cleaning* kemudian menghapus karakter yang tidak diperlukan, seperti tautan URI, tanda baca, simbol, emotikon, dan angka, untuk menyisakan elemen alfabetis murni [12]. Normalisasi leksikon juga diterapkan guna mengonversi istilah gaul (*slang*) ke bentuk formal.

Pasca keseragaman format, mekanisme *tokenizing* memecah untaian kalimat menjadi token kata individual. Analisis kemudian dipertajam lewat *stopword removal*, yang berfungsi menghapus kata yang kurang relevan dan minim makna. Kendati demikian, intervensi ini secara spesifik mempertahankan leksikon negasi krusial (misalnya: "belum", "tidak", "bukan") lantaran perannya yang sangat vital dalam membalikkan orientasi sentimen opini. Tahapan pembersihan ini berperan dalam kelayakan data dan berkorelasi langsung dengan ketajaman prediksi model [13].

2.5 Transformasi Fitur dan Segmentasi Data

Rangkaian matematis dan prosedural diberlakukan guna mematangkan struktur informasi sebelum diolah oleh algoritma *machine learning*. Adapun mekanismenya terbagi sebagai berikut :

a. Pemberian Label Data

Anotasi korpus diformulasikan ke dalam trikotomi sentimen (positif, netral, negatif) yang dijangkarkan pada perolehan skor bintang pengguna [14]. Pendekatan pelabelan berbasis *rating* dipandang sangat representatif, mengingat skor kuantitatif tersebut merupakan proksi valid atas derajat kepuasan konsumen. Metodologi ini lumrah diadaptasi dalam studi penambahan opini literatur terkemuka [15]. Secara teknis, skor 1 dan 2 diposisikan sebagai kelas negatif, *rating* 3 sebagai pandangan netral, sedangkan skor 4 dan 5 mewakili sentimen positif [16].

b. Pemisahan Data (*Data Splitting*)

Sebelum pemodelan, korpus dibagi menjadi *training set* dan *testing set* bersandarkan rasio 80:20. Penyekatan ini dieksekusi dengan teknik *stratified split* untuk mempertahankan proporsi setiap kelas pada data latih dan data uji, sehingga distribusi kelas pada kedua bagian *dataset* tetap terjaga [19].

c. Matriks *Term Frequency-Inverse Document Frequency* (TF-IDF)

Leksikon tekstual wajib dikonversi menjadi larik numerik sebelum tahapan komputasi berjalan. Metode TF-IDF digunakan untuk memberikan bobot pada setiap *term* berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat kemunculannya dalam seluruh korpus [17].

$$TF-IDF(t, d) = (1 + \log f_{t,d}) \times \log \left(\frac{N}{df(t)} \right) \quad (1)$$

Konstruksi matematis pada Persamaan (1) dirancang sedemikian rupa agar *Term Frequency* berkonfigurasi logaritmik mampu menjinakkan hiper-dominasi kata yang muncul berulang di satu dokumen. Sementara itu, Komponen *Inverse Document Frequency* memberikan bobot lebih tinggi pada *term* yang muncul pada lebih sedikit dokumen dan bobot lebih rendah pada *term* yang muncul pada banyak dokumen. Hasil akhirnya, leksikon umum akan menerima penalti bobot, sedangkan kata kunci esensial akan terekstraksi optimal [18].

2.6 Penanganan Ketidakseimbangan Kelas

Hasil pengolahan data menunjukkan adanya ketidakseimbangan distribusi kelas, terutama pada sentimen netral yang memiliki jumlah sampel jauh lebih rendah dibandingkan kelas positif dan negatif. Pada data pengujian yang terdiri atas 2.723 observasi, kelas netral berjumlah 69 data. Kondisi tersebut dapat menyebabkan model lebih berorientasi pada kelas mayoritas sehingga kemampuan mengenali kelas minoritas menjadi terbatas. Untuk mengurangi pengaruh ketidakseimbangan tersebut, penelitian menerapkan parameter *class_weight='balanced'* pada *Logistic Regression* dan *Support Vector Machine*. Konfigurasi tersebut memberikan bobot lebih besar terhadap kelas dengan jumlah sampel lebih sedikit sehingga kesalahan klasifikasi pada kelas minoritas memperoleh penalti yang lebih tinggi. Namun, mekanisme ini tidak menambah jumlah sampel secara fisik sehingga keterbatasan data pada kelas netral tetap perlu diperhatikan dalam interpretasi hasil. Oleh karena itu, evaluasi model tidak hanya didasarkan pada *accuracy*, tetapi juga menggunakan *precision*, *recall*, *F1-score*, *macro recall*, *macro F1-score*, dan *confusion matrix* untuk mengevaluasi kemampuan model pada setiap kelas.

2.7 Pemodelan dan Analisis Kinerja

a. Desain Pemodelan

Tahap ini membandingkan *Logistic Regression* (LR) dan SVM menggunakan matriks TF-IDF. Kedua metode ini dipilih berdasarkan rekam jejak keandalannya saat berhadapan dengan data tekstual berdimensi tinggi [20]. Patut dicatat bahwa penelitian ini difokuskan pada pengujian kinerja dasar (*baseline performance*). Oleh karena itu, optimasi *hyperparameter* menggunakan metode pencarian seperti *GridSearchCV* tidak diterapkan. Kedua model menggunakan konfigurasi parameter yang telah ditetapkan sebelumnya dan diterapkan secara konsisten pada *dataset*, *preprocessing*, dan skema pembagian data yang sama. Pendekatan ini digunakan agar perbandingan lebih berfokus pada perbedaan karakteristik klasifikasi kedua algoritma, bukan pada perbedaan tingkat optimasi *hyperparameter*.

b. Implementasi Algoritma *Logistic Regression* (LR)

LR bertindak sebagai klasifikator probabilitas linier berbasis fungsi sigmoid (Persamaan 2). Pemodelan algoritma ini dikonfigurasi dengan penalti reguler dan secara eksplisit menggunakan parameter *solver='liblinear'*. Pemilihan *solver liblinear* didasarkan pada kesesuaiannya untuk *dataset* klasifikasi linear berdimensi tinggi, dukungan terhadap regularisasi L1 dan L2, serta kemampuan menangani pendekatan multikelas *One-vs-Rest* (OvR) pada konfigurasi tiga kelas. Selain itu, penggunaan *solver* ini diposisikan sebagai *baseline* penelitian, bukan untuk tujuan optimasi *solver* [21]. Dalam penelitian ini, konfigurasi tersebut diterapkan bersama *class_weight='balanced'* untuk memberikan bobot yang lebih besar terhadap kelas dengan jumlah sampel lebih sedikit. Formulasi fungsi probabilitas logistik didefinisikan sebagai:

$$P(Y = k | X) = \frac{e^{\beta_k^T X}}{\sum_{j=1}^K e^{\beta_j^T X}} \quad (2)$$

c. Implementasi Algoritma *Support Vector Machine* (SVM)

Kerja komputasi SVM dikonsentrasikan pada pencarian batas pemisah (*hyperplane*) yang optimal yang diapit oleh ukuran margin terjauh. Taktik ini memungkinkan algoritma memisahkan data pada ruang fitur berdimensi tinggi secara presisi dan tegas [22]. Hal tersebut menjadi salah satu karakteristik SVM dalam klasifikasi data teks di ranah analisis sentimen [23]. Persamaan dasar SVM direpresentasikan sebagai :

$$\hat{y} = \arg \max_{k \in \{\text{positif, netral, negatif}\}} (w_k^T x + b_k) \quad (3)$$

d. Pengujian Kinerja Model

Tahap evaluasi diaplikasikan guna mengukur tingkat reliabilitas algoritma saat memprediksi kategori sentimen secara akurat. Pengukuran performa ini difasilitasi sepenuhnya oleh instrumen *confusion matrix* (matriks konfusi).

1. *Accuracy*

Parameter akurasi merepresentasikan metrik evaluasi yang menghitung persentase ketepatan prediksi model terhadap label kelas yang sesungguhnya (baik negatif, netral, maupun positif). Pada konteks studi ini, rasio *accuracy* mendeskripsikan kapabilitas komprehensif dari algoritma untuk memilah keseluruhan data ulasan GuitarTuna secara tepat sasaran [24].

2. *Precision*

Metrik *precision* mengkalkulasi proporsi atau rasio antara data yang diprediksi positif dengan benar oleh sistem dibandingkan dengan total keseluruhan data yang diprediksi masuk ke dalam kelas tersebut. Perolehan skor presisi yang masif mengindikasikan bahwa model memiliki tingkat keandalan yang baik dalam menebak kelas sentimen yang aktual [25].

3. *Recall*

Sebagai indikator krusial dalam evaluasi performa model prediktif, *recall* difungsikan untuk mengkalkulasi rasio keberhasilan mesin dalam menemukan kembali seluruh sampel aktual yang berafiliasi dengan kelas sentimen tertentu. Bersandar pada landasan literatur komputasi mutakhir yang dikemukakan oleh Zhang dkk. [26], instrumen ini secara spesifik merepresentasikan proporsi instans target di dalam korpus data yang sukses diidentifikasi dengan tepat oleh algoritma tanpa terlewat. Oleh sebab itu, eskalasi perolehan skor pada metrik recall menjadi bukti empiris yang mengisyaratkan bahwa sistem klasifikasi beroperasi dengan tingkat recall yang relatif tinggi. Kondisi ini menunjukkan bahwa sebagian besar ulasan dapat dikenali oleh model dari proses pemindaian dan berhasil didistribusikan secara presisi ke dalam kompartemen yang semestinya.

4. *F1-Score*

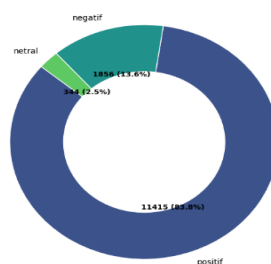
F1-Score dirumuskan melalui kalkulasi rata-rata harmonis yang menggabungkan elemen *precision* serta recall menjadi satu kesatuan skor komprehensif. Indikator ini menawarkan pandangan evaluasi yang jauh lebih proporsional, terkhusus saat algoritma harus menghadapi kompleksitas identifikasi untuk opini-opini berkelas netral di dalam himpunan data ulasan GuitarTuna [27].

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengolahan Data

Agregasi awal menghasilkan 20.339 dokumen teks dari Google Play Store, App Store, serta kuesioner pengguna GuitarTuna. Selanjutnya, dilakukan proses *data cleaning* untuk mengeliminasi data duplikat, teks kosong, dan respons yang tidak relevan sehingga diperoleh 13.615 data bersih yang digunakan dalam penelitian. Distribusi sentimen pada 13.615 data setelah proses *cleaning* ditunjukkan pada Gambar 2.

Distribusi Sentimen Pengguna (Positif, Netral, Negatif)



Gambar 2. Distribusi Sentimen

Pada Gambar 2, terlihat distribusi pelabelan setelah proses *cleaning* menunjukkan ketimpangan proporsi yang signifikan. Kelas positif terdiri atas 11.415 data (83,8%), kelas negatif sebanyak 1.856 data (13,6%), dan kelas netral sebanyak 344 data (2,5%). Dengan demikian, jumlah keseluruhan pada distribusi tersebut adalah 13.615 data yang digunakan sebagai *dataset* untuk tahap pemodelan. Ketimpangan ekstrem ini mengindikasikan bahwa opini audiens

cenderung condong pada kepuasan, namun menimbulkan tantangan tersendiri karena model berisiko mengalami bias dengan mengabaikan kelas minoritas.

Guna meredam dampak ketimpangan tersebut, kerangka pelatihan mengadopsi parameter *class_weight='balanced'*. Konfigurasi tersebut memberikan bobot lebih besar terhadap kelas dengan jumlah sampel lebih sedikit sehingga kesalahan pada kelas minoritas memperoleh penalti lebih tinggi. Kendati demikian, injeksi penalti ini tidak mereplikasi jumlah data sampel netral secara fisik, sehingga defisit volume pada kelas minoritas tersebut tetap menjadi catatan penting saat menginterpretasikan tingkat keandalan hasil prediksi akhir.

3.2 Hasil Pre-processing

Fase purifikasi teks dieksekusi guna merapikan struktur data mentah sebelum dikonversi menjadi matriks numerik. Rangkaian ini mencakup *case folding*, *cleaning*, normalisasi *slang*, *stopword removal*, hingga *tokenization*. Ilustrasi transformasi pada tiap tahapan dirangkum melalui Tabel 3.

Tabel 3. Hasil *Pre-processing*

Alur Proses	Hasil
Teks Asli	Mantap bisa <i>offline</i> juga Hasil setelan cocok 👍
<i>Case Folding</i>	mantap bisa <i>offline</i> juga hasil setelan cocok 👍
<i>Cleaning</i>	mantap bisa <i>offline</i> juga hasil setelan cocok
<i>Stopword Removal</i>	mantap <i>offline</i> hasil setelan cocok
<i>Tokenization</i>	mantap <i>offline</i> hasil setelan cocok

Merujuk pada Tabel 3, setiap tahapan modifikasi sukses menyeterilkan bentuk leksikon tanpa merusak substansi informasi yang merefleksikan pengalaman riil pengguna. Penyeragaman huruf lewat *case folding* disusul pemusnahan emotikon serta karakter asing pada *cleaning* berhasil membersihkan deretan kata. Reduksi kata tugas via *stopword removal* menyisakan leksikon kunci yang esensial, yang kemudian dipecah menjadi token individual agar siap dipetakan dalam ekstraksi fitur TF-IDF.

3.3 Hasil Pembobotan TF-IDF

Algoritma TF-IDF berfungsi mengonversi korpus tekstual yang telah dibersihkan menjadi representasi vektor numerik. Perolehan pembobotan menunjukkan bahwa kata-kata dengan bobot TF-IDF tertinggi berkaitan dengan karakteristik penggunaan GitarTuna, seperti *guitar* dan *tuning*. Di samping itu, leksikon evaluatif seperti *good*, bagus, *great*, *best*, dan *easy* mencerminkan apresiasi kualitas serta fungsionalitas kemudahan navigasi. Munculnya istilah asing seperti *buena* dan *es* juga menunjukkan adanya penggunaan variasi bahasa di luar bahasa Indonesia pada ulasan pengguna. Kendati demikian, tingginya nilai bobot TF-IDF tidak serta-merta mengindikasikan bahwa suatu kata merupakan representasi mutlak dari sentimen positif atau negatif, melainkan menunjukkan derajat signifikansi kemunculan *term* dalam dokumen dan korpus. Oleh sebab itu, pembacaan orientasi sentimen tetap dikaji secara terpisah melalui mekanisme klasifikasi serta pemetaan kata spesifik per kelas.

3.4 Hasil Logistic Regression

Kinerja algoritma *Logistic Regression* dalam memisahkan opini audiens ke dalam tiga kategori (positif, negatif, serta netral) dapat diamati melalui hasil klasifikasi yang diperoleh. Guna memetakan efektivitas model secara komprehensif, proses evaluasi mengandalkan instrumen pengukuran berupa *accuracy*, *precision*, *recall*, beserta *F1-score*. Di samping mengevaluasi skor agregat, peninjauan juga difokuskan pada capaian spesifik di tiap kategori sentimen, mengingat adanya potensi ketidakseimbangan proporsi pada *dataset*. Detail visualisasi dari luaran klasifikasi ini dapat disimak pada representasi berikut.

```

=== Logistic Regression ===
              precision    recall  f1-score   support

negatif      0.78        0.60        0.68        371
netral       0.30        0.10        0.15         69
positif      0.92        0.98        0.95       2283

accuracy          0.90        0.90        0.90       2723
macro avg       0.67        0.56        0.59       2723
weighted avg    0.89        0.90        0.89       2723
  
```

Akurasi Logistic Regression: 90.30%

Gambar 3. Hasil Klasifikasi *Logistic Regression*

Gambar 3 menjelaskan bahwa model *Logistic Regression* diimplementasikan untuk memetakan ulasan ke dalam trikotomi polaritas (positif, negatif, dan netral). Hasil pengujian menunjukkan tingkat akurasi global (*accuracy*) sebesar 90,30%. Sekalipun angka tersebut tampak superior secara akumulatif, evaluasi terperinci per kelas memperlihatkan sebaran performa yang tidak merata.

Pada kelas positif, model ini mampu mengenali sebagian besar sampel pada kelas tersebut, terbukti dari perolehan *recall* 0,98 dan *F1-score* 0,95. Tingginya performa pada kelas positif berkaitan dengan jumlah data positif yang lebih besar dibandingkan kelas lainnya, sehingga model memiliki lebih banyak contoh untuk mempelajari karakteristik bahasa pada kelas tersebut. Sebaliknya, performa pada kelas negatif lebih rendah, dengan *recall* 0,60 dan *F1-score* 0,68. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam mengenali sebagian data negatif.

Kendala paling kentara terkonsentrasi pada kelas netral, di mana model mencatatkan *precision* 0,30, *recall* 0,10, dan *F1-score* 0,15. Rendahnya rasio *recall* menunjukkan bahwa sebagian besar opini netral gagal terdeteksi oleh sistem. Hal ini sekali lagi dipicu oleh minimnya populasi sampel netral, sehingga struktur tata bahasanya sulit dipelajari secara komprehensif. Temuan ini menegaskan bahwa metrik akurasi 90,30% harus dievaluasi secara bersisian dengan metrik per kelas agar performa model tidak bias oleh dominasi kelas positif.

3.5 Hasil Support Vector Machine

Luaran dari pemodelan SVM diaplikasikan untuk meninjau kapabilitas algoritma saat memetakan opini Guitartuna ke dalam tiga spektrum sentimen. Proses peninjauan efektivitas ini bertumpu pada indikator *precision*, *recall*, *F1-score*, beserta *accuracy*. Tiap-tiap metrik evaluasi tersebut menyajikan sudut pandang yang unik terkait keandalan sistem dalam mengidentifikasi kategori sentimen. Rincian luaran evaluasi algoritma SVM dapat dicermati melalui representasi visual di bawah ini.

```

=== SVM ===

```

	precision	recall	f1-score	support
negatif	0.76	0.71	0.73	371
netral	0.22	0.14	0.17	69
positif	0.94	0.96	0.95	2283
accuracy			0.91	2723
macro avg	0.64	0.60	0.62	2723
weighted avg	0.90	0.91	0.90	2723

Akurasi SVM: 90.53%

Gambar 4. Hasil Klasifikasi Support Vector Machine

Gambar 4 menunjukkan bahwa algoritma SVM membukukan tingkat akurasi (*accuracy*) sebesar 90,53% pada data uji. Perolehan ini hanya terpaut selisih tipis 0,23 poin persentase dari *Logistic Regression*. Oleh karena itu, analisis performa tidak bertumpu semata pada angka akurasi global, melainkan ditelisik melalui dinamika pada masing-masing kelas sentimen.

Pada ranah positif, SVM mempertahankan performa tinggi dengan *F1-score* 0,95. Sementara pada kategori negatif, SVM menghasilkan *recall* 0,71 dan *F1-score* 0,73, angka yang lebih unggul dibandingkan capaian *Logistic Regression*. Lonjakan ini merefleksikan kapabilitas SVM dalam menangkap lebih banyak sampel negatif pada data uji. Kendati demikian, tren positif tersebut belum merata hingga ke kategori netral.

Kelas netral tetap menjadi titik terlemah bagi model SVM. Nilai *precision* sebesar 0,22, *recall* 0,14, dan *F1-score* 0,17 mengindikasikan bahwa model masih menghadapi kendala serius dalam membedakan ulasan netral dari kelompok lainnya. Keterbatasan kuantitas data uji kelas netral (sebanyak 69 entri) menyebabkan informasi pendukung sangat minim. Meskipun parameter *class_weight='balanced'* telah diaktifkan untuk memihak kelas minoritas, intervensi tersebut belum mampu menutupi defisit pasokan data latih yang memadai untuk mengenali variasi bahasa netral.

3.6 Perbandingan Kinerja Model

Komparasi performa antara *Logistic Regression* dan SVM dievaluasi di atas basis data pengujian yang seragam. Secara matematis, *Logistic Regression* menghasilkan akurasi 90,30%, sedangkan SVM mengantongi 90,53% (selisih 0,23 poin persentase). Selisih tersebut menunjukkan bahwa perbandingan berdasarkan *accuracy* saja belum cukup untuk menyimpulkan adanya perbedaan kinerja yang berarti antara kedua model, yang dapat dilihat pada Tabel 4 berikut.

Tabel 4. Perbandingan Kinerja Logistic Regression dan SVM

Model	Accuracy	Macro Recall	Macro F1-Score
Logistic Regression	90,30%	0,56	0,59
SVM	90,53%	0,60	0,62

Merujuk pada ringkasan Tabel 4, SVM unggul secara tipis dalam hal perolehan *macro recall* dan *macro F1-score*. Perbedaan tersebut terutama terlihat pada kategori negatif, sementara kedua algoritma sama-sama memiliki performa rendah pada kelas netral. Hal ini menunjukkan bahwa evaluasi model perlu mempertimbangkan performa

pada setiap kelas, terutama kelas dengan jumlah sampel yang lebih sedikit, bukan sekadar melihat angka akurasi keseluruhan.

Guna menelusuri kecenderungan kesalahan klasifikasi secara lebih mendetail, prosedur pengujian diteruskan dengan mengeksploitasi matriks konfusi (*confusion matrix*) pada masing-masing algoritma. *Confusion matrix* digunakan untuk menunjukkan jumlah data yang diklasifikasikan dengan benar maupun yang mengalami *misclassification* pada setiap kelas.

3.6.1 Uji Signifikansi McNemar

Untuk menguji apakah terdapat perbedaan kinerja yang signifikan antara *Logistic Regression* dan *Support Vector Machine*, penelitian menerapkan uji *McNemar* pada data pengujian yang sama. Pemilihan uji ini didasarkan pada karakteristik prediksi kedua model yang berasal dari observasi berpasangan. *Uji McNemar* berfokus pada pasangan diskordan, yaitu kondisi ketika salah satu model menghasilkan prediksi yang benar sedangkan model lainnya menghasilkan prediksi yang salah. Kerangka hipotesis dalam pengujian ini diformulasikan sebagai berikut :

H_0 : Tidak terdapat perbedaan kinerja klasifikasi yang signifikan secara statistik antara *Logistic Regression* dan *Support Vector Machine*.

H_1 : Terdapat perbedaan kinerja klasifikasi yang signifikan secara statistik antara *Logistic Regression* dan *Support Vector Machine*.

Untuk mengetahui apakah terdapat perbedaan kinerja yang signifikan antara *Logistic Regression* dan *Support Vector Machine*, dilakukan uji *McNemar*. Hasil tabulasi prediksi kedua model ditunjukkan pada Tabel 5.

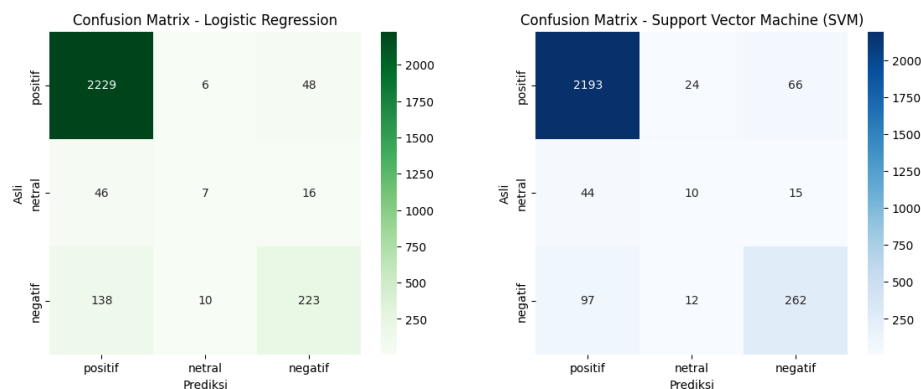
Tabel 5. Kontingensi McNemar

	SVM Benar	SVM Salah
LR Benar	2407	52
LR Salah	58	206

Berdasarkan Tabel 5, terdapat 2.407 data yang diklasifikasikan dengan benar oleh kedua model dan 206 data yang sama-sama diklasifikasikan salah. Sementara itu, terdapat 52 data yang hanya diklasifikasikan dengan benar oleh *Logistic Regression* dan 58 data yang hanya diklasifikasikan dengan benar oleh *Support Vector Machine*. Dua nilai tersebut merupakan pasangan data *discordant* yang menjadi dasar dalam uji *McNemar*. Pada *exact McNemar test*, statistik uji diperoleh dari nilai terkecil antara jumlah data yang hanya benar pada *Logistic Regression* dan jumlah data yang hanya benar pada *Support Vector Machine*, sehingga diperoleh nilai statistik $\min(52, 58) = 52$. Selanjutnya, *p-value* dihitung berdasarkan distribusi binomial dengan jumlah pasangan *discordant* sebesar $52 + 58 = 110$ dan probabilitas 0,5, yang menghasilkan *p-value* sebesar 0,6338. Dengan menggunakan taraf signifikansi $\alpha = 0,05$, nilai *p-value* sebesar 0,6338 lebih besar daripada 0,05. Dengan demikian, tidak terdapat bukti yang cukup untuk menyatakan adanya perbedaan kinerja yang signifikan secara statistik antara *Logistic Regression* dan *Support Vector Machine* pada data pengujian.

3.7 Analisis Confusion Matrix

Untuk melihat distribusi prediksi benar dan salah pada setiap kelas sentimen, evaluasi dilanjutkan menggunakan *confusion matrix* pada kedua model. Hasil perbandingan *confusion matrix Logistic Regression* dan *Support Vector Machine* ditampilkan pada Gambar 5.



Gambar 5. *Confusion Matrix Logistic Regression dan Support Vector Machine*

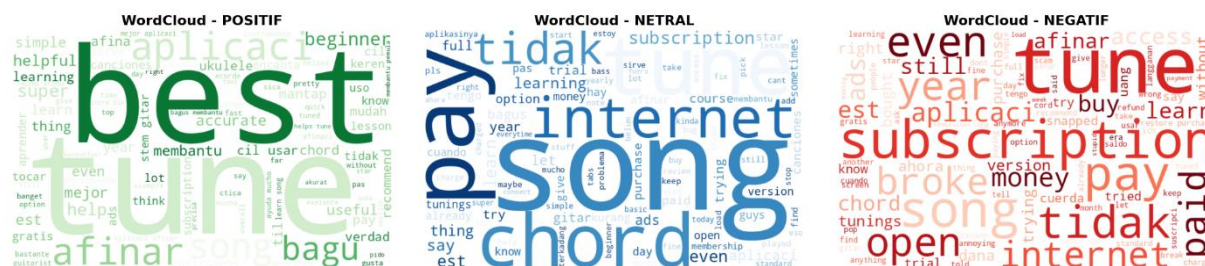
Berdasarkan matriks konfusi *Logistic Regression* pada gambar 5 di atas, kesalahan tebakan terbanyak bermuara pada pergeseran sampel kelas negatif dan netral yang salah diprediksi sebagai sentimen positif. Hal ini menunjukkan bahwa sebagian data dari kelas negatif dan netral masih salah diprediksi sebagai kelas positif, yang sejalan dengan rendahnya recall pada kedua kelas tersebut.

Pola dislokasi serupa juga terdeteksi pada model SVM di mana migrasi data negatif-netral masih mengarah ke kelas positif. Akan tetapi, secara kuantitatif, SVM memperoleh recall yang lebih tinggi pada kelas negatif dan netral dibandingkan *Logistic Regression*, meskipun performa pada kelas netral masih rendah. Peningkatan ketepatan ini terproyeksi langsung pada perolehan recall dan *F1-score* kelas negatif dan netral yang lebih tinggi, meskipun ruang gerak pada kelas netral masih terbatas oleh minimnya volume sampel.

Secara komprehensif, matriks konfusi mengonfirmasi bahwa kendala utama riset ini bukan terletak pada penanganan kelas positif, melainkan pada pemisahan kelompok minoritas (khususnya netral) dari dominasi kelas positif dan negatif. Temuan ini memvalidasi urgensi metrik *macro F1-score* bahwa tingginya angka akurasi belum merefleksikan pemerataan performa model secara adil di seluruh kelas.

3.8 Analisis Kata Dominan Berdasarkan Sentimen

Analisis kata dominan disematkan untuk memperkaya interpretasi isi ulasan pada masing-masing kategori polaritas. Berbeda dari TF-IDF yang bertindak sebagai vektor matematis pemodelan, analisis ini berfokus memetakan esensi topik yang paling sering disuarakan pengguna. Visualisasi gabungan disajikan melalui Gambar berikut.



Gambar 6. WordCloud Sentimen Positif, Netral, dan Negatif

Pada gambar 6, terlihat bahwa pada kluster positif, kemunculan leksikon seperti *best*, *tune*, *helpful*, *useful*, dan *simple* merefleksikan kepuasan audiens terhadap *utilitas tuning*, kemudahan operasional, serta manfaat aplikasi. Adanya istilah *learning*, *lesson*, dan *chord* juga mengindikasikan apresiasi terhadap modul edukasi gitar. Hal ini menunjukkan bahwa pengalaman positif pengguna tidak terbatas pada fitur penyyetaman semata, melainkan merambah pada aspek pembelajaran interaktif.

Pada kluster netral, dominasi kata seperti *song*, *tune*, *pay*, *chord*, *internet*, *subscription*, dan *trial* menunjukkan bahwa opini netral berkutat seputar fungsionalitas dan batas akses fitur. Namun, leksikon ini tidak bisa serta-merta diartikan sebagai sikap objektif murni atau pertimbangan transaksi, mengingat label netral dalam penelitian ini dikunci secara mutlak berdasarkan *rating* 3, bukan hasil anotasi semantik manual.

Pada kluster negatif, kemunculan istilah seperti *tune*, *subscription*, *pay*, *internet*, *access*, *paid*, dan *broke* menyingkap problem krusial terkait limitasi fitur, hambatan transaksi pembayaran, isu koneksi, hingga kendala operasional. Temuan ini memberikan konteks tambahan terhadap hasil klasifikasi, terutama pada tema yang berkaitan dengan fitur berbayar, koneksi, dan akses aplikasi. Dengan demikian, analisis kata dominan berfungsi efektif sebagai instrumen pelengkap pembacaan karakteristik data.

3.9 Pembahasan

Hasil penelitian menunjukkan bahwa TF-IDF dapat digunakan untuk merepresentasikan korpus ulasan GuitarTuna dalam bentuk numerik yang kemudian digunakan oleh *Logistic Regression* dan SVM. Kedua model sukses mencetak akurasi global melampaui 90%, namun distribusinya terpantau timpang pada setiap kelas sentimen. Kategori positif mendulang pengenalan terbaik berkat limpahan volume data, sementara kategori netral mencatatkan performa terendah akibat kelangkaan sampel. Kondisi tersebut menunjukkan bahwa distribusi kelas perlu dipertimbangkan dalam menginterpretasikan hasil evaluasi model.

Hasil riset ini menunjukkan koherensi dengan temuan Budianto dkk. [7] yang mengomparasikan *Logistic Regression* dan SVM pada ulasan aplikasi e-commerce Android, di mana SVM membukukan performa yang sedikit superior. Perbedaananya terletak pada ranah penerapan, di mana studi ini mengeksplorasi domain aplikasi musik digital. Perbedaan domain menunjukkan bahwa karakteristik *dataset* dan leksikon dapat menjadi salah satu faktor yang perlu dipertimbangkan dalam menginterpretasikan perbedaan performa algoritma. Pada korpus GuitarTuna, tantangan fundamental berpusat pada fragmentasi kelas netral yang sangat terbatas.

Studi ini juga memperluas spektrum investigasi yang dirintis oleh Ariq [2] mengenai analisis ulasan GuitarTuna dan Fender Guitar Tuner menggunakan *Naïve Bayes* dan SVM. Penelitian terdahulu belum mengeksplorasi komparasi antara SVM dan *Logistic Regression* serta cakupan datanya terbatas pada platform tunggal. Melalui integrasi data hibrida (Google Play Store, App Store, dan kuesioner primer), riset ini menyajikan perspektif komparatif yang lebih holistik mengenai efektivitas algoritma linier pada ranah utilitas musik digital.

Penerapan ekstraksi TF-IDF dalam penelitian ini sejalan dengan kerangka kerja Lestari dan Hutagalung [5] pada ulasan layanan logistik. Kendati demikian, karakteristik leksikon pada korpus GuitarTuna memiliki keunikan

tersendiri yang sarat akan istilah teknis musik, penyyetaman, dan kosakata multibahasa, yang menjadi variabel pembeda signifikansi hasil antar-penelitian.

Di sisi lain, temuan ini memperkuat konklusi Maulana dkk. [8] serta Cahyani dan Prasetyaningrum [9] bahwa evaluasi model tidak sebaiknya hanya didasarkan pada satu metrik. SVM memperoleh *accuracy* sebesar 90,53%, sedangkan *Logistic Regression* memperoleh 90,30%. SVM juga memiliki *macro recall* dan *macro F1-score* yang lebih tinggi. Namun, hasil uji McNemar menghasilkan *p-value* sebesar 0,6338 sehingga belum terdapat bukti yang cukup untuk menunjukkan adanya perbedaan kinerja yang signifikan secara statistik antara kedua model. Dengan demikian, selisih *accuracy* sebesar 0,23 poin persentase belum dapat digunakan sebagai dasar untuk menyatakan adanya perbedaan performa yang signifikan antara *Logistic Regression* dan SVM pada *dataset* penelitian ini.

Secara keseluruhan, kedua algoritma mampu mencapai *accuracy* di atas 90%, tetapi performanya belum merata pada seluruh kelas sentimen. Kelas netral menjadi kelas dengan performa terendah pada kedua model karena jumlah sampelnya terbatas. Penerapan *class_weight*='balanced' membantu memberikan bobot lebih besar terhadap kelas minoritas, tetapi tidak menambah jumlah sampel yang tersedia untuk dipelajari oleh model. Oleh karena itu, evaluasi perlu mempertimbangkan *accuracy*, metrik per kelas, *macro recall*, *macro F1-score*, *confusion matrix*, serta hasil uji signifikansi McNemar agar perbandingan kinerja kedua model dapat diinterpretasikan secara lebih menyeluruh.

4. KESIMPULAN

Penelitian ini mengevaluasi kinerja algoritma *Logistic Regression* dan *Support Vector Machine* (SVM) yang dikombinasikan dengan ekstraksi fitur TF-IDF untuk menganalisis sentimen pengguna aplikasi GuitarTuna. Pada tahap persiapan, proses prapemrosesan berhasil mereduksi 20.339 data mentah menjadi 13.615 ulasan bersih. Langkah penyaringan ini difokuskan untuk mengeliminasi *noise* seperti komentar kosong, karakter tidak relevan, dan teks duplikat. Meskipun penyaringan diperlukan untuk menjaga kualitas pemodelan, penyusutan volume data tetap memiliki kemungkinan menimbulkan bias seleksi terhadap data yang dipertahankan. Secara evaluasi kuantitatif, algoritma SVM memperoleh akurasi sebesar 90,53%, *macro recall* 0,60, dan *macro F1-score* 0,62, sedangkan *Logistic Regression* memperoleh akurasi 90,30%, *macro recall* 0,56, serta *macro F1-score* 0,59. Meskipun SVM memiliki nilai metrik yang sedikit lebih tinggi, hasil uji McNemar menunjukkan *p-value* sebesar 0,6338, sehingga perbedaan performa kedua model tidak signifikan secara statistik. Analisis klasifikasi juga menunjukkan adanya ketimpangan kinerja akibat distribusi kelas yang tidak merata. Kedua algoritma mampu mengklasifikasikan kelas positif dengan baik, tetapi masih mengalami kesulitan dalam mengidentifikasi kelas netral karena jumlah data latih yang terbatas. Kondisi ini menunjukkan bahwa penggunaan *accuracy* saja belum cukup untuk menggambarkan kinerja model pada seluruh kelas. Oleh karena itu, *macro recall*, *macro F1-score*, dan *confusion matrix* diperlukan untuk memberikan gambaran performa yang lebih menyeluruh. Dari perspektif praktis, sentimen positif sebagian besar berkaitan dengan keandalan fitur penyyetaman alat musik dan kemudahan antarmuka pengguna, sedangkan sentimen negatif banyak berkaitan dengan biaya langganan, keterbatasan fitur, dan kendala teknis transaksi. Temuan tersebut dapat menjadi informasi bagi pengembang dalam mengevaluasi aspek layanan dan fitur yang berkaitan dengan pengalaman pengguna. Secara akademis, penelitian ini memberikan gambaran empiris mengenai penerapan *Logistic Regression* dan SVM berbasis TF-IDF pada klasifikasi sentimen ulasan aplikasi GuitarTuna, khususnya pada kondisi data dengan distribusi kelas yang tidak seimbang. Dengan demikian, hasil penelitian tidak hanya memberikan perbandingan kinerja kedua algoritma, tetapi juga menunjukkan pentingnya mempertimbangkan karakteristik distribusi kelas dalam mengevaluasi model klasifikasi sentimen. Sebagai tindak lanjut, penelitian selanjutnya disarankan untuk memperbanyak sampel pada kelas minoritas, menerapkan teknik penanganan imbalanced data, melakukan optimasi hyperparameter, serta mengeksplorasi penggunaan metode *deep learning* untuk membandingkan kinerja pada karakteristik data yang sama.

5. PERNYATAAN PENULIS

5.1 Kontribusi Penulis

Konseptualisasi: M. Y., dan M. R. S.;

Metodologi: M. Y., dan M. R. S.;

Perangkat Lunak: M. Y.;

Validasi: M. Y., dan M. R. S.;

Analisis Formal: M. Y., dan M. R. S.;

Investigasi: M. Y., dan M. R. S.;

Sumber Daya: M. Y., dan M. R. S.;

Kurasi Data: Muhammad Yusuf;

Penulisan Draf Awal: M. Y., dan M. R. S.;

Penulisan, Peninjauan, dan Penyuntingan: M. Y., M. R. S., D. R. I., dan A. A.;

Visualisasi: M. Y., dan M. R. S.;

Seluruh penulis telah membaca dan menyetujui versi manuskrip yang telah diterbitkan.

5.2 Ketersediaan Data

Data yang disajikan dalam penelitian ini tersedia berdasarkan permintaan kepada penulis korespondensi.

5.3 Pendanaan

Penelitian ini tidak menerima pendanaan eksternal.

5.4 Pernyataan Dewan Peninjau Institusional

Tidak Berlaku (Not applicable).

5.5 Pernyataan Persetujuan Berdasarkan Informasi

Tidak Berlaku (Not applicable).

5.6 Pernyataan Konflik Kepentingan

Para penulis menyatakan bahwa mereka tidak memiliki kepentingan keuangan yang bersaing atau hubungan pribadi yang diketahui yang dapat dianggap memengaruhi pekerjaan yang dilaporkan dalam artikel ini.

5.7 Penggunaan AI Generatif dan Teknologi Berbantuan AI dalam Proses Penulisan

Para penulis menggunakan AI Generatif untuk meningkatkan kejelasan penulisan dalam artikel ini. Para penulis telah meninjau dan menyunting konten yang dihasilkan dengan bantuan AI serta bertanggung jawab penuh atas versi akhir publikasi.

REFERENCES

- [1] I. R. Nugroho, I. Wilujeng, and Z. K. Prasetyo, "Mobile Learning Research Trends in Indonesia: A Systematic Literature Review," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 8, pp. 356–365, 2023, doi: 10.29303/jppipa.v9i8.3073.
- [2] A. Ariq, "Analisis Sentimen Positif Aplikasi GitarTuna dan Fender Guitar Tuner dengan Menggunakan Metode Algoritma Naïve Bayes dan Support Vector Machine," Universitas Pembangunan Nasional "Veteran" Jakarta, Jakarta, 2024. Accessed: Apr. 27, 2026. [Online]. Available: <http://repository.upnvj.ac.id/id/eprint/28184>
- [3] A. Gour, S. Aggarwal, and S. Kumar, "Lending Ears to Unheard Voices: An Empirical Analysis of User-Generated Content on Social Media," *Prod. Oper. Manag.*, vol. 31, no. 6, pp. 2457–2476, 2022, doi: 10.1111/poms.13732.
- [4] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *TEKNOSAINS : Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, Jul. 2023, doi: 10.37373/tekno.v10i2.419.
- [5] V. B. Lestari and C. A. Hutagalung, "Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," *J-KOMA: Jurnal Ilmu Komputer dan Aplikasi*, vol. 8, no. 1, pp. 36–44, 2025, doi: 10.21009/j-koma.v8i1.05.
- [6] H. P. Singh, N. Singh, A. Mishra, S. K. Sen, M. Swarnkar, and D. Pandey, "Logistic Regression Based Sentiment Analysis System: Rectify," in *2024 IEEE International Conference on Big Data & Machine Learning (ICBDML)*, IEEE, 2024, pp. 186–191. doi: 10.1109/ICBDML60909.2024.10577296.
- [7] A. G. Budianto, R. Rusilawati, A. T. E. Suryo, G. R. Cahyono, A. F. Zulkarnain, and M. Martunus, "Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android," *Jurnal Sains dan Informatika*, vol. 10, no. 2, 2024, doi: 10.34128/jsi.v10i2.911.
- [8] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, Feb. 2024, doi: 10.57152/malcom.v4i2.1206.
- [9] R. D. Cahyani and P. T. Prasetyaningrum, "Sentiment Analysis of User Reviews for AI Applications: Evaluating SVM, Logistic Regression, and Random Forest," *Journal of Information Systems and Informatics*, vol. 8, no. 1, pp. 1–27, 2026, doi: 10.63158/journalisi.v8i1.1366.
- [10] U. Sulung and M. Muspawi, "Memahami Sumber Data Penelitian: Primer, Sekunder, dan Tersier," *Jurnal Edu Research Indonesian Institute For Corporate Learning And Studies (IICLS)*, 2024, doi: 10.47827/jer.v5i3.238.
- [11] Yuyun, A. D. Latief, T. Sampurno, Hazriani, A. O. Arisha, and Mushaf, "Next Sentence Prediction: The Impact of Preprocessing Techniques in Deep Learning," in *Proceedings - 2023 10th International Conference on Computer, Control, Informatics and its Applications: Exploring the Power of Data: Leveraging Information to Drive Digital Innovation, IC3INA 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 274–278. doi: 10.1109/IC3INA60834.2023.10285805.
- [12] M. A. Setiawan, M. H. Efendi, M. F. Akbar, and W. S. Pratama, "Pengaruh Teknik Preprocessing terhadap Kinerja Model Explainable Boosting Machine (EBM) untuk Prediksi Serangan Jantung," *Nusantara of Engineering (NOE)*, vol. 8, no. 02, pp. 317–326, 2025, doi: 10.29407/noe.v8i02.25811.
- [13] B. Liu, "Opinion Spam Detection," in *Sentiment analysis and opinion mining*, Springer, 2022, pp. 113–125. Accessed: Apr. 22, 2026. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-02145-9_10
- [14] S. Mujilawati, "Analisis Sentimen Pengguna Aplikasi ChatGPT Berdasarkan Rating Menggunakan Metode Lexicon," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 9, no. 1, pp. 131–137, Dec. 2023, doi: 10.36341/rabit.v9i1.3845.
- [15] K. Ferhati, A. Burlea-Schiopoiu, and A.-G. Nascu, "A Text-Based Project Risk Classification System Using Multi-Model AI: Comparing SVM, Logistic Regression, Random Forests, Naive Bayes, and XGBoost," *Systems*, vol. 13, no. 12, p. 1078, 2025, doi: 10.3390/systems13121078.

- [16] G. Darmawan, S. Alam, M. Imam Sulisty, P. Studi Teknik Informatika, S. Tinggi Teknologi Wastukencana Purwakarta, and R. Artikel, “Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi MyPertamina pada Google Playstore Menggunakan Metode Naïve Bayes,” *STORAGE – Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 2, no. 3, pp. 100–108, 2023, doi: 10.55123/storage.v2i3.2333.
- [17] M. S. Amrullah, A. G. Putrada, M. N. Fauzan, and N. Alamsyah, “ETLE Sentiment Analysis Performance Increase with TF-IDF, MDI Feature Selection, and SVM,” *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 4, pp. 1308–1318, 2024, doi: 10.32520/stmsi.v13i4.2701.
- [18] S. A. Sazan, M. H. Miraz, and A. B. M. Rahman, “Enhancing Depressive Post Detection in Bangla: A Comparative Study of TF-IDF, BERT and FastText Embeddings,” *arXiv preprint arXiv:2407.09187*, 2024, doi: 10.48550/arXiv.2407.09187.
- [19] H. A. Taher, “Dual-Language Sentiment Analysis: A Comprehensive Evaluating SVM, Logistic Regression, XGBoost, and Decision Tree Using TF-IDF On Arabic and English Dataset,” *Wasit Journal for Pure sciences*, vol. 3, no. 4, pp. 59–69, 2024, doi: 10.31185/wjps.549.
- [20] Nida Nur Aini Aryanti and Ozzi Suria, “Analisis Sentimen terhadap Pemutusan Hubungan Kerja di Indonesia: Komparasi IndoBERT dengan SVM, Random Forest, dan Decision Tree dengan Optimasi TF-IDF,” *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 1158–1176, Jul. 2025, doi: 10.36341/rabit.v10i2.6364.
- [21] A. J. N. Siagian and E. D. Tarigan, “Model Regresi Logistik Multinomial untuk Mengetahui Pengaruh Faktor-Faktor Penggunaan Kacamata Minus pada Pelajar SMA Sutomo 1,” in *SEMIOTIKA (Seminar Nasional Teknologi Informasi dan Matematika)*, 2023, pp. 58–76. Accessed: Apr. 23, 2026. [Online]. Available: <https://journal.itk.ac.id/index.php/semiotika/article/view/998>
- [22] S. K. Papia, M. A. Khan, T. Habib, M. Rahman, and M. N. Islam, “DistilRoBiLSTMFuse: An Efficient Hybrid Deep Learning Approach for Sentiment Analysis,” *PeerJ Comput. Sci.*, vol. 10, p. e2349, 2024, doi: 10.7717/peerj-cs.2349.
- [23] A. A. Firdaus, A. Yudhana, and I. Riadi, “Analisis Sentimen Pada Proyeksi Pemilihan Presiden 2024 Menggunakan Metode Support Vector Machine,” *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 236–245, 2023, doi: 10.51454/decode.v3i2.172.
- [24] A. Sultana, M. Hasan, M. Shidujaman, and C. Premachandra, “Sentiment Analysis with Deep Learning Methods for Performance Assessment and Comparison,” in *2024 International Conference on Image Processing and Robotics (ICIPRoB)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ICIPRoB62548.2024.10544219.
- [25] F. H. Hilmi and A. D. Indriyanti, “Sentiment Analysis of 2024 Election Fraud Using SVM and Naïve Bayes Algorithms,” *Journal of Emerging Information Systems and Business Intelligence*, vol. 5, no. 4, pp. 353–364, 2024, doi: 10.26740/jeisbi.v5i4.64703.
- [26] A. Zhang, Z. C. Lipton, M. U. Li, and A. J. Smola, *Dive into Deep Learning*. Cambridge University Press, 2023. Accessed: Apr. 27, 2026. [Online]. Available: https://books.google.co.id/books?hl=id&lr=&id=vfDiEAAAQBAJ&oi=fnd&pg=PR17&dq=A.+Zhang,+Z.+C.+Lipton,+M.+U.+Li,+and+A.+J.+Smola,+%E2%80%9CDive+into+Deep+Learning,%E2%80%9D+2023.&ots=ZQkaDuwAnj&sig=j8EC8D_kXIC_5YNShygic0yeeQ&redir_esc=y#v=onepage&q&f=false
- [27] Y. Xu, B. Klein, G. Li, and B. Gopaluni, “Evaluation of Logistic Regression and Support Vector Machine Approaches for XRF Based Particle Sorting for a Copper Ore,” *Miner. Eng.*, vol. 192, p. 108003, 2023, doi: 10.1016/j.mineng.2023.108003.