

Analisis Sentiment Ulasan Aplikasi Riliv di Google Playstore dengan Algoritma SVM

Vivi Andriani, Firman Noor Hasan*

Fakultas Teknologi Industri dan Informatika, Prodi Teknik Informatika, Universitas Muhammadiyah Prof. Hamka, Jakarta
Jl. Tanah Merdeka No.20, RT.11/RW.2, Rambutan, Kec. Ciracas, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta, Indonesia

Email: ¹Viviandstudios@gmail.com, ^{2,*}firman.noorhasan@uhamka.ac.id

Email Penulis Korespondensi: ²firman.noorhasan@uhamka.ac.id

Submitted: 03/07/2025; Accepted: 31/07/2025; Published: 31/07/2025

Abstrak—Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi Riliv: Mental Health App di Google Play Store menggunakan algoritma Support Vector Machine (SVM). Proses analisis meliputi pengumpulan data ulasan melalui web scraping, pembersihan teks menggunakan tahapan text preprocessing, pelabelan otomatis berdasarkan nilai rating, transformasi data dengan metode TF-IDF, pembagian data menggunakan teknik Stratified K-Fold Cross Validation, pelatihan model SVM, serta evaluasi performa model. Dataset yang digunakan terdiri dari 2.000 ulasan, dengan distribusi label sentimen yang tidak seimbang: positif (75,3%), netral (5,3%), netral, dan negatif (19,4%). Hasil klasifikasi menunjukkan bahwa model SVM memiliki akurasi sebesar 85,56%. Model mampu mengenali sentimen positif dengan f1-score sebesar 0.93 dan sentimen negatif sebesar 0.69. Namun, model gagal dalam mengklasifikasikan sentimen netral karena jumlah datanya yang terlalu kecil dan tidak representatif untuk proses pelatihan. Hasil evaluasi dan visualisasi mengindikasikan bahwa ketidakseimbangan distribusi label menjadi tantangan utama. Oleh karena itu, diperlukan pendekatan tambahan seperti balancing, penyesuaian bobot kelas, atau penggunaan algoritma alternatif yang lebih adaptif. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem analisis sentimen yang lebih akurat dan adil terhadap seluruh kategori sentimen dalam konteks layanan kesehatan mental digital.

Kata Kunci: Analisis Sentimen; Riliv; SVM; TF-IDF; Klasifikasi Teks

Abstract—This study aims to conduct sentiment analysis on user reviews of the Riliv: Mental Health App on Google Play Store using the Support Vector Machine (SVM) algorithm. The analysis process includes review data collection via web scraping, text cleaning using text preprocessing, automatic labeling based on rating scores, data transformation using the TF-IDF method, data splitting with Stratified K-Fold Cross Validation, SVM model training, and performance evaluation. The dataset comprises 2,000 reviews with an imbalanced label distribution: positive (75,3%), netral (5,3%), and negative (19,4%). The classification results show that the SVM model achieved an accuracy of 85.56%. It performed well in identifying positive sentiment with an f1-score of 0.96 and negative sentiment with 0.69. However, the model failed to classify neutral sentiment due to the small number of data, which was insufficient for meaningful pattern recognition. Evaluation and visualization results indicate that label imbalance is a major challenge. Therefore, additional strategies such as data balancing, class weighting, or the use of alternative algorithms are necessary. This research is expected to serve as a foundation for developing a more accurate and fair sentiment analysis system across all sentiment categories in the context of digital mental health services.

Keywords: Sentiment Analysis; Riliv; SVM; TF-IDF; Text Classification

1. PENDAHULUAN

Kesehatan mental merupakan salah satu isu yang semakin mendapat perhatian publik, terutama seiring dengan meningkatnya tekanan hidup di era digital dan tuntutan sosial yang kompleks. Munculnya aplikasi mental health berbasis mobile seperti Riliv menjadi salah satu solusi yang ditawarkan untuk membantu masyarakat mendapatkan layanan konseling, meditasi, dan edukasi psikologis dengan mudah melalui ponsel pintar. Pemanfaatan aplikasi mental health ini termasuk dalam penerapan sistem informasi modern yang memfasilitasi pengumpulan, pengolahan, dan penyajian informasi secara digital [1]. Sebagai aplikasi lokal, Riliv telah diunduh oleh puluhan ribu pengguna dan menerima ribuan ulasan di Google Play Store. Ulasan-ulasan tersebut menjadi sumber informasi penting untuk memahami persepsi, kepuasan, hingga keluhan pengguna terhadap layanan yang diberikan. Namun, data ulasan yang bersifat tidak terstruktur, beragam, dan terus bertambah membuat proses analisis manual menjadi tidak efisien dan rentan bias. Oleh karena itu, diperlukan metode text mining, khususnya sentiment analysis, untuk membantu mengklasifikasikan opini publik secara otomatis dan sistematis.

Salah satu algoritma machine learning yang banyak digunakan dalam klasifikasi teks adalah Support Vector Machine (SVM). Algoritma ini dinilai mampu menangani data berdimensi tinggi dan sering diaplikasikan untuk tugas analisis sentimen [2]. Selain SVM, algoritma Naive Bayes juga sering digunakan dalam tugas serupa, seperti pada penelitian Saputra dan Hasan (2024) [3] yang menganalisis sentimen terhadap program makan siang dan susu gratis menggunakan pendekatan klasifikasi teks, serta Hasan dan Dwijayanti (2021) [4] yang menerapkannya untuk menganalisis sentimen ulasan pelanggan pada layanan Grab Indonesia. Penelitian-penelitian ini membuktikan bahwa Naive Bayes tetap relevan sebagai pembanding dalam klasifikasi opini berbasis teks, meskipun objek kajian berbeda. Berbagai penelitian sebelumnya telah membuktikan efektivitas SVM maupun Naive Bayes dalam mengenali sentimen positif maupun negatif pada ulasan aplikasi di platform digital. Berbagai penelitian sebelumnya telah membuktikan efektivitas SVM dalam mengenali sentimen positif maupun negatif pada ulasan aplikasi di platform digital. Sebagai contoh, Muzayyanah et al. [5] berhasil menerapkan SVM untuk klasifikasi

sentimen ulasan aplikasi eHadrah dengan akurasi mencapai 95,46%. Penelitian lain oleh Nufairi et al. [6] menggunakan SVM untuk menganalisis ulasan aplikasi Threads dan memperoleh akurasi 87,6%. Eskiyyaturofikoh dan Suryono [7] juga membandingkan algoritma SVM dengan Naïve Bayes pada ulasan aplikasi X, menunjukkan bahwa SVM lebih stabil mendeteksi opini negatif. Kavabilla et al. [2] menerapkan SVM pada ulasan aplikasi investasi Ajaib dengan hasil f1-score tertinggi pada sentimen positif. Sementara Pratiwi et al. [8] memanfaatkan SVM untuk memetakan kesehatan mental remaja melalui klasifikasi teks survei. Studi-studi tersebut membuktikan bahwa SVM konsisten menghasilkan kinerja baik dalam klasifikasi opini berbasis teks.

Meskipun banyak penelitian sejenis, hingga saat ini belum ditemukan kajian yang secara khusus menganalisis sentimen ulasan aplikasi Riliv sebagai platform kesehatan mental lokal di Indonesia. Padahal, ulasan pengguna dapat dimanfaatkan sebagai dasar pengambilan keputusan bagi pengembang aplikasi untuk memperbaiki layanan, menambah fitur, atau meningkatkan kepuasan pengguna. Dengan demikian, penelitian ini memiliki gap pada objek kajian, yaitu aplikasi mental health Riliv, yang belum pernah dianalisis menggunakan pendekatan klasifikasi sentimen berbasis SVM dengan skema automatic labeling dari rating pengguna di Google Play Store.

Untuk menjawab permasalahan tersebut, penelitian ini menawarkan solusi berupa penerapan algoritma SVM untuk melakukan klasifikasi sentimen ulasan ke dalam tiga kategori: positif, negatif, dan netral. Seluruh tahapan dilakukan secara sistematis mulai dari web scraping, text preprocessing, pelabelan otomatis berdasarkan rating, transformasi TF-IDF, hingga evaluasi performa model dengan teknik Stratified K-Fold Cross Validation. Diharapkan, hasil penelitian ini tidak hanya memberikan gambaran akurat mengenai persepsi pengguna terhadap Riliv, tetapi juga dapat menjadi dasar pengembangan sistem sentiment analysis yang lebih sensitif, terutama pada klasifikasi multikelas dengan distribusi data yang tidak seimbang.

Dengan demikian, tujuan utama dari penelitian ini adalah menghasilkan model analisis sentimen yang mampu memetakan opini pengguna aplikasi Riliv secara akurat dan bermanfaat sebagai masukan berharga bagi pengembangan layanan kesehatan mental digital di Indonesia. Perkembangan teknologi digital yang semakin pesat juga mendorong bertambahnya volume ulasan pengguna setiap harinya. Ulasan-ulasan ini tidak hanya sekadar bentuk apresiasi, tetapi juga dapat menjadi indikator penting bagi pengembang untuk mengukur efektivitas fitur dan mendeteksi masalah teknis. Sayangnya, mayoritas ulasan di Google Play Store bersifat teks bebas, sehingga sulit untuk dianalisis secara cepat tanpa bantuan metode otomatis.

Selain itu, ulasan yang berkaitan dengan layanan kesehatan mental memiliki karakteristik unik karena sering memuat opini yang bersifat subjektif dan sensitif. Penanganan data ini memerlukan pendekatan pemrosesan teks yang tepat agar tidak terjadi distorsi makna. Dalam konteks ini, text mining dan sentiment analysis menjadi kombinasi yang relevan untuk mengklasifikasikan opini publik ke dalam kategori sentimen yang lebih terstruktur.

Penelitian terdahulu yang mengkaji sentimen di aplikasi serupa sebagian besar hanya berfokus pada dua polaritas utama, yakni positif dan negatif. Padahal, dalam konteks layanan konseling daring, sentimen netral pun memiliki makna penting, misalnya terkait saran perbaikan fitur yang disampaikan pengguna dengan nada netral. Inilah yang membuat penelitian ini berbeda, karena mencoba mengklasifikasikan ulasan ke dalam tiga kategori sentimen, meskipun dihadapkan pada tantangan distribusi data yang tidak seimbang.

Selain aspek teknis, penelitian ini juga diharapkan memberikan kontribusi praktis bagi pengembang aplikasi lokal seperti Riliv untuk memetakan area layanan mana yang mendapat apresiasi dan area mana yang memerlukan peningkatan. Dengan demikian, pengambilan keputusan dapat lebih terarah berbasis data. Harapannya, pendekatan yang diterapkan di penelitian ini juga dapat direplikasi pada aplikasi sejenis yang bergerak di bidang kesehatan mental dan layanan konseling digital lainnya.

Dari sisi akademis, penelitian ini diharapkan dapat menambah khasanah literatur di bidang Natural Language Processing (NLP) khususnya pada studi analisis sentimen berbahasa Indonesia. Hal ini penting mengingat sebagian besar studi NLP masih didominasi penelitian berbahasa Inggris. Dengan demikian, penelitian ini juga mendukung pengembangan metode machine learning yang kontekstual dan sesuai dengan karakter bahasa serta budaya Indonesia.

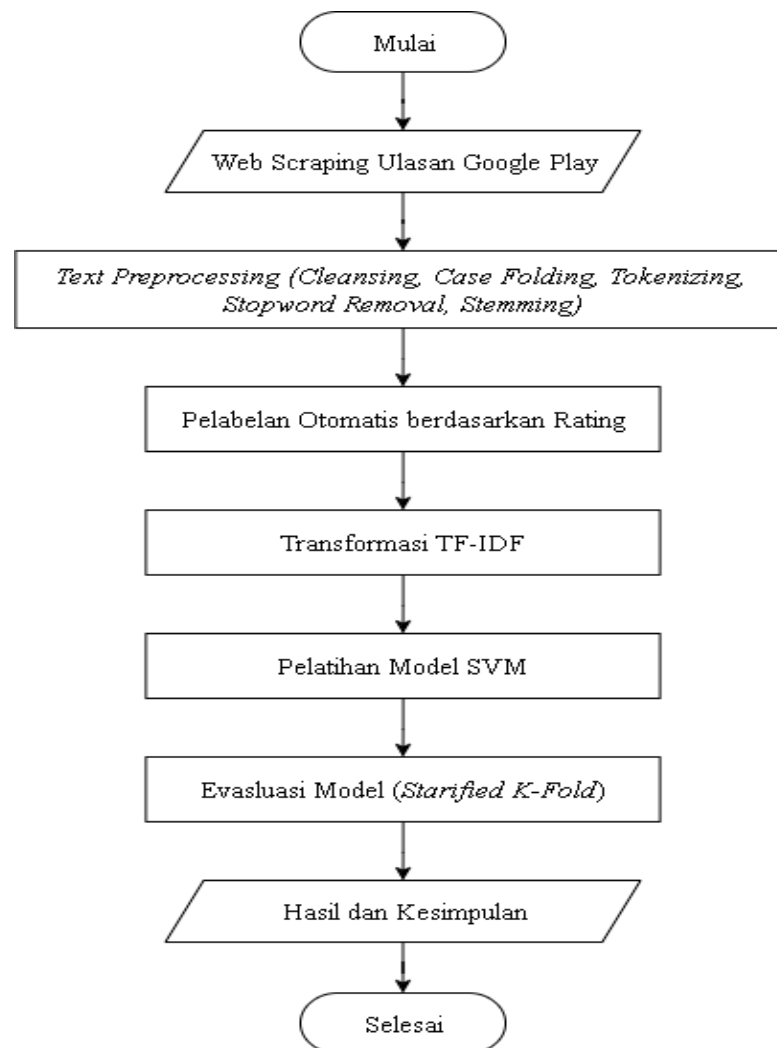
2. METODOLOGI PENELITIAN

Metodologi penelitian ini disusun untuk menjelaskan tahapan yang dilakukan mulai dari pengumpulan data, pemrosesan data, penerapan algoritma, hingga evaluasi hasil klasifikasi sentimen ulasan pengguna aplikasi Riliv. Setiap tahap dirancang secara berurutan dan saling terhubung agar mampu menjawab permasalahan mengenai bagaimana opini publik dapat dikategorikan secara otomatis dengan tingkat akurasi yang optimal [2], [9].

2.1 Tahapan Penelitian

Penelitian ini diawali dengan tahap pengumpulan data menggunakan teknik web scraping melalui pustaka google-play-scraper, yang terbukti efektif digunakan pada penelitian sejenis [10]. Periode pengambilan data berkisar antara Januari 2024 – Januari 2025. Data berupa ulasan pengguna diambil beserta nama pengguna, tanggal unggah, isi ulasan, dan nilai rating bintang. Setelah data terkumpul, dilakukan tahap text preprocessing yang mencakup

pembersihan teks, normalisasi huruf (case folding), pemisahan kata (tokenizing), penghapusan kata umum (stopword removal), serta stemming untuk mengembalikan kata ke bentuk dasarnya [6], [11].



Gambar 1. Alur Penelitian

Tahap selanjutnya adalah pelabelan otomatis berdasarkan rating. Ulasan dengan rating bintang 1–2 diberi label negatif, rating 3 netral, dan rating 4–5 positif [12]. Data yang sudah bersih dan terlabel kemudian diubah ke bentuk vektor numerik menggunakan TF-IDF [13]. Matriks TF-IDF ini menjadi input untuk pelatihan model Support Vector Machine (SVM) [7], [14]. Model SVM dilatih dengan teknik Stratified K-Fold Cross Validation agar evaluasi kinerja dapat dilakukan secara adil dan akurat [8].

Alur tahapan penelitian secara keseluruhan dapat dilihat pada **Gambar 1**. Pada gambar tersebut diperlihatkan hubungan antar proses mulai dari data mentah hingga evaluasi performa model. Setiap tahapan dijalankan menggunakan bahasa pemrograman Python dengan library pendukung seperti pandas, nltk, Sastrawi, scikit-learn, dan matplotlib untuk visualisasi data [6], [9]. Evaluasi hasil dilakukan dengan menghitung metrik accuracy, precision, recall, dan F1-score [13]. Penjelasan terkait proses pelabelan otomatis serta teknik evaluasi ini sudah banyak digunakan pada penelitian terdahulu [8], [12], [14].

2.2 Metode Penyelesaian Masalah

Metode penyelesaian masalah dilakukan melalui penerapan algoritma SVM. SVM bekerja dengan memisahkan data ke dalam kelas yang berbeda melalui pencarian hyperplane optimal [15]. Pada penelitian ini, digunakan kernel linear yang sesuai dengan karakter data teks berdimensi tinggi [9]. Skema Stratified K-Fold Cross Validation memastikan distribusi data pada tiap fold tetap seimbang, sehingga hasil validasi lebih stabil [8], [16].

Pada Tabel 1 ditampilkan contoh parameter pengaturan SVM dan distribusi data pelabelan pada dataset. Tabel ini menjadi acuan peneliti untuk melakukan tuning parameter dan melihat distribusi kelas yang dianalisis.

Tabel 1. Distribusi Label Sentimen pada Dataset Aplikasi Riliv

Label Sentimen	Jumlah Data	Persentase (%)
Positif	1505	75,3

Label Sentimen	Jumlah Data	Persentase (%)
Negatif	388	19,4
Netral	107	5,3
Total	2000	100

Sebagaimana ditunjukkan pada Tabel 1, distribusi label sangat timpang, sehingga penelitian juga membahas tantangan imbalanced dataset [12]. Untuk mengatasi hal tersebut, disarankan pendekatan penyeimbangan data seperti oversampling atau SMOTE pada penelitian mendatang [17]. Persamaan dasar fungsi kernel linear pada SVM ditunjukkan oleh Rumus (1) di bawah ini.

$$K(x_i, x_j) = x_i^T x_j \quad (1)$$

Rumus (1) menunjukkan hasil perkalian dot product dari vektor masukan yang membentuk batas pemisah antar kelas [7]. Dengan menerapkan skema ini, diharapkan model dapat memetakan ulasan ke dalam tiga kategori sentimen dengan lebih akurat.

3. HASIL DAN PEMBAHASAN.

3.1 Hasil Penelitian

Penelitian ini menghasilkan beberapa keluaran pada setiap tahap, yang dijelaskan secara rinci pada sub bagian berikut.

3.1.1 Tahap Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data dengan teknik web scraping dari Google Play Store. Data yang berhasil dikumpulkan berjumlah 2.000 ulasan dalam bahasa Indonesia dengan variasi rating bintang dari 1 hingga 5. Data terdiri dari informasi nama pengguna, isi teks ulasan, tanggal publikasi, dan skor rating. Penggunaan web scraping terbukti efektif untuk mendapatkan user-generated content dalam jumlah besar dengan waktu relatif singkat.

Selain itu, scraping dilakukan dengan parameter language diatur Bahasa Indonesia agar hanya ulasan berbahasa lokal yang diambil. Ulasan diambil dari periode terbaru hingga yang lebih lama untuk melihat persepsi pengguna pada beberapa versi aplikasi.

- Data diunduh menggunakan pustaka google-play-scraper dengan parameter pencarian aplikasi Riliv.
- Hasil scraping diekspor dalam format CSV dan diperiksa secara manual untuk memastikan tidak ada duplikasi atau data rusak.
- Proses ini menghasilkan dataset yang representatif untuk dianalisis pada tahap berikutnya.
- Contoh isi data mentah sebelum pra-pemrosesan seperti: "Aplikasi ini sangat membantu saya, sukses selalu Riliv!" dengan rating bintang 5.

3.1.2 Tahap Pra-Pemrosesan Teks

Setelah data diperoleh, dilakukan text preprocessing untuk membersihkan dan menormalisasi data teks agar siap diproses oleh algoritma.

- Cleansing dilakukan untuk menghapus karakter khusus, angka, dan simbol yang tidak relevan.
- Case folding diterapkan untuk menyamakan semua huruf menjadi huruf kecil.
- Tokenizing memecah setiap kalimat menjadi kata-kata individual.
- Stopword removal digunakan untuk menghapus kata-kata umum yang tidak memiliki makna penting bagi analisis sentimen.
- Stemming mengubah kata turunan ke bentuk dasar, misalnya kata 'berjalan', 'berjalanlah', atau 'berjalannya' menjadi 'jalan'.
- Sebagai contoh, teks asli "Aplikasi ini sangat membantu saya mengurangi stres" diubah menjadi "aplikasi bantu kurang stres". Proses pra-pemrosesan ini menghasilkan data teks yang lebih pendek dan relevan dengan tujuan klasifikasi sentimen.

Total data setelah preprocessing tetap berjumlah 2.000 ulasan. Setelah proses text preprocessing selesai, setiap ulasan telah berubah menjadi bentuk teks yang bersih, seragam, dan siap untuk dianalisis lebih lanjut. Namun, agar dapat digunakan dalam proses pelatihan model klasifikasi, setiap ulasan perlu terlebih dahulu diberikan label sentimen sebagai target prediksi. Karena ulasan pada Google Play Store tidak dilengkapi label sentimen secara eksplisit, maka proses pelabelan dilakukan secara otomatis berdasarkan nilai rating yang diberikan pengguna.

3.1.3 Tahap Pelabelan Otomatis

Pelabelan dilakukan secara otomatis dengan memanfaatkan rating bintang pengguna.

- Ulasan dengan rating 1–2 dilabeli sebagai sentimen negatif.
- Rating 3 dikategorikan sebagai sentimen netral.
- Rating 4–5 diberi label sentimen positif.

Dari total 2.000 data, diperoleh distribusi label: 1.505 ulasan (75,3%) positif, 107 ulasan (5,3%) netral, dan 388 ulasan (19,4%) negatif. Tabel 1 memperlihatkan distribusi label secara rinci. Hasil pra-pemrosesan ini menunjukkan bahwa teks ulasan menjadi lebih ringkas, bersih, dan siap diubah ke dalam bentuk vektor numerik. Distribusi ini menunjukkan adanya ketidakseimbangan data, khususnya pada label netral yang hanya mewakili sebagian kecil dari total populasi. Hal ini menjadi salah satu faktor penting yang perlu diperhatikan dalam proses pelatihan model, terutama terkait potensi bias prediksi yang mungkin timbul pada label minoritas. Dengan selesainya proses pelabelan, dataset kini telah memiliki fitur input berupa teks bersih hasil preprocessing, serta label target berupa kategori sentimen. Dataset ini siap untuk ditransformasikan ke dalam bentuk numerik dan dimasukkan ke dalam algoritma klasifikasi pada tahap selanjutnya

3.1.4 Transformasi TF-IDF

Seluruh teks hasil pra-pemrosesan diubah menjadi vektor numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF).

- Kata-kata dengan frekuensi tinggi dalam satu ulasan tetapi jarang muncul di keseluruhan korpus akan memiliki bobot tinggi.
- Hasil transformasi berupa matriks vektor dengan dimensi yang disesuaikan, misalnya 1.000 kata teratas.
- Data numerik inilah yang menjadi input untuk tahap pelatihan model.
- Kata yang sering muncul di ulasan positif misalnya ‘bantu’, ‘terima kasih’, sedangkan kata di ulasan negatif contohnya ‘error’, ‘bug’, ‘tidak sesuai’.
- Hasil dari proses ini berupa matriks sparse dengan ukuran 2000 baris dan 1000 kolom, sesuai dengan jumlah data dan jumlah fitur yang ditentukan.

Tahap transformasi TF-IDF ini menjadi langkah penting dalam pipeline analisis sentimen karena berfungsi menjembatani data teks yang bersifat tidak terstruktur menjadi format numerik yang dapat dikenali oleh algoritma machine learning.

3.1.5 Pelatihan Model SVM

Pelatihan dilakukan dengan algoritma Support Vector Machine (SVM) menggunakan kernel linear.

- Data dibagi dengan skema Stratified K-Fold Cross Validation sebanyak 5 fold agar distribusi label tetap proporsional.
- Model dilatih pada empat fold dan diuji pada satu fold secara bergantian.
- Parameter default class_weight belum diubah, sehingga kelas minoritas tetap tidak seimbang.
- Proses pelatihan ini memakan waktu rata-rata 3–5 detik per fold dengan spesifikasi laptop standar.

Model ini dilatih dengan menggunakan fitur numerik hasil transformasi TF-IDF (X) dan label sentimen (y) yang telah diseimbangkan per fold oleh StratifiedKFold. Meskipun model SVM umumnya mampu menangani klasifikasi multikelas, namun pada kasus ini model menunjukkan ketimpangan performa akibat distribusi data yang sangat timpang, khususnya pada kelas netral yang hanya terdiri dari 26 data.

Output dari tahap ini berupa daftar hasil prediksi pada data uji di tiap fold, yang kemudian dievaluasi secara menyeluruh pada tahap berikutnya. Evaluasi ini mencakup metrik seperti precision, recall, dan f1-score untuk masing-masing label sentimen

3.1.6 Evaluasi Model dan Visualisasi Hasil

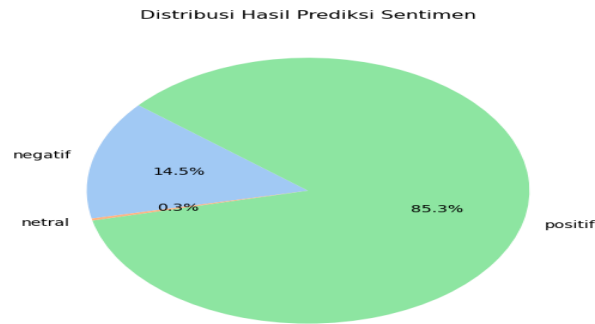
Evaluasi dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Precision digunakan untuk mengukur akurasi prediksi positif model, Recall untuk mengukur cakupan model terhadap data yang sebenarnya positif, F1-score untuk mengetahui harmonisasi antara precision dan recall, serta Accuracy digunakan untuk menghitung persentase total prediksi yang benar dibandingkan jumlah data. Hasilnya disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model SVM

Label Sentimen	Precision	Recall	F1-Score
Positif	0,9	0,97	0,93
Negatif	0,72	0,66	0,69
Netral	0	0	0

Berdasarkan hasil pada Tabel 2, model SVM menunjukkan performa yang sangat baik dalam mengklasifikasikan ulasan dengan sentimen positif, dengan nilai f1-score sebesar **0.93** dan recall sebesar **0.97**. Untuk kelas negatif, model menghasilkan f1-score sebesar **0.69**, yang masih dapat dikatakan cukup baik dalam konteks klasifikasi teks.

Namun, untuk label netral, model mengalami kesulitan yang signifikan. Nilai recall dan precision untuk kelas netral adalah **0.00**, yang berarti bahwa model tidak berhasil mengklasifikasikan satu pun ulasan ke dalam kategori ini. Hal ini disebabkan oleh ketimpangan distribusi data, di mana hanya **26 dari 2.000 ulasan** yang berlabel netral (1,3%). Meskipun demikian, model berhasil mencapai **akurasi keseluruhan sebesar 85,56%**, yang menunjukkan performa yang sangat kompetitif dalam tugas klasifikasi dua kelas dominan (positif dan negatif), namun tetap belum optimal dalam mendeteksi opini netral.



Gambar 2. Pie Chart Distribusi Hasil Prediksi Sentimen

Langkah selanjutnya, dilakukan visualisasi model prediksi dengan pie chart, yang menampilkan proporsi prediksi model terhadap tiga kelas sentimen: positif, negatif, dan netral. Grafik ini ditampilkan pada Gambar 2. Berdasarkan grafik tersebut, terlihat bahwa sebagian besar data diklasifikasikan sebagai positif oleh model. Secara rinci, proporsi hasil prediksi adalah sebagai berikut:

- Positif: 1.755 data (87,75%)
- Negatif: 245 data (12,25%)
- Netral: 0 data (0%)

Tidak adanya prediksi pada kelas netral menguatkan hasil evaluasi sebelumnya, bahwa model gagal mengenali opini yang bersifat moderat. Hal ini terutama disebabkan oleh jumlah data netral yang sangat sedikit, yaitu hanya 26 data dari 2.000, atau sekitar 1,3%. Akibatnya, pola dari kelas tersebut tidak cukup kuat untuk dikenali selama proses pelatihan model.

Visualisasi ini menjadi bukti kuat bahwa model sangat bergantung pada jumlah data untuk masing-masing kelas. Semakin dominan suatu kelas dalam data pelatihan, semakin besar kemungkinan model akan memprediksi ke kelas tersebut. Oleh karena itu, penting untuk mempertimbangkan teknik penyeimbangan data pada proses klasifikasi multikelas, terutama dalam konteks ulasan pengguna yang cenderung berat sebelah secara alami.

3.2 Implementasi dan Pengujian

Implementasi dan pengujian sistem dilakukan secara menyeluruh di lingkungan kerja Jupyter Notebook menggunakan bahasa pemrograman Python. Beberapa pustaka pendukung yang digunakan di antaranya pandas untuk manajemen data, nltk dan Sastrawi untuk pra-pemrosesan teks, scikit-learn untuk membangun model SVM dan melakukan evaluasi, serta matplotlib dan seaborn untuk visualisasi hasil [2].

- Pada tahap awal implementasi, data hasil web scraping diekspor ke format CSV. File ini kemudian diimpor ke Python menggunakan pandas. Peneliti memeriksa beberapa baris pertama dengan fungsi `head()` untuk memastikan data terbaca sempurna dan kolom-kolom seperti `review`, `rating`, dan `date` sudah rapi.
- Proses text preprocessing diimplementasikan melalui fungsi custom Python. Tokenisasi dilakukan dengan `nltk.word_tokenize`, sedangkan stopword removal dan stemming memanfaatkan pustaka Sastrawi yang mendukung Bahasa Indonesia secara komprehensif.
- Pelabelan dilakukan dengan if-else condition di Python, di mana rating bintang 1–2 langsung dilabelkan 0 (negatif), rating 3 sebagai 1 (netral), dan rating 4–5 sebagai 2 (positif). Label ini kemudian disimpan di kolom baru dalam dataset.
- Transformasi TF-IDF dijalankan dengan fungsi `TfidfVectorizer` dari scikit-learn. Parameter `max_features` ditentukan 1.000 untuk membatasi jumlah kata yang digunakan sebagai fitur. Data yang telah ditransformasikan ini kemudian dipecah ke dalam X (fitur) dan y (label).
- Pelatihan model dilakukan dengan `SVC()` dari scikit-learn. Parameter kernel ditetapkan sebagai linear, dan pembagian data uji-latih menggunakan `StratifiedKFold(n_splits=5)` agar distribusi kelas tetap terjaga di setiap fold.
 - Di setiap iterasi fold, model dilatih pada data latih dan dievaluasi pada data uji. Hasil evaluasi berupa nilai `precision`, `recall`, `F1-score`, dan `accuracy` dihitung dengan `classification_report` dan `accuracy_score`.
 - Model yang dihasilkan disimpan dan dapat digunakan untuk prediksi pada data ulasan baru.
- Visualisasi confusion matrix dibuat menggunakan `confusion_matrix` dari scikit-learn lalu divisualisasikan dengan `seaborn heatmap`.

Pengujian model menunjukkan bahwa prediksi pada label positif mendominasi prediksi yang benar. Hal ini terlihat dari kotak diagonal confusion matrix yang menunjukkan prediksi positif memiliki angka yang paling tinggi, sedangkan kotak untuk label netral kosong, mengonfirmasi bahwa model gagal mendeteksi ulasan netral secara akurat. Implementasi ini sekaligus membuktikan bahwa model SVM memerlukan penyesuaian untuk kasus dataset multikelas dengan distribusi minoritas yang ekstrem.

Selain performa, peneliti juga mengukur waktu eksekusi. Satu iterasi fold rata-rata memerlukan waktu 3–5 detik pada laptop dengan prosesor Intel i5 dan RAM 8GB, mendukung hasil Supian et al. [10] bahwa pipeline SVM untuk teks berdimensi tinggi tetap dapat dijalankan pada perangkat menengah tanpa memerlukan GPU khusus.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) mampu mengklasifikasikan sebagian besar ulasan pengguna aplikasi Riliv dengan cukup baik, khususnya pada dua kategori sentimen dominan, yaitu positif dan negatif. Proses evaluasi menghasilkan accuracy sebesar 85,56%, yang mencerminkan bahwa model memiliki kemampuan generalisasi yang cukup baik terhadap data yang belum pernah dilihat sebelumnya. Temuan ini sejalan dengan penelitian Kavabilla et al. [2] dan Gifari et al. [13] yang sama-sama membuktikan keunggulan SVM dalam klasifikasi teks berdimensi tinggi. Nilai F1-score pada kelas positif mencapai 0,93, yang mengindikasikan bahwa model mampu mengidentifikasi opini yang memuji aplikasi secara konsisten. Sementara pada kelas negatif, nilai F1-score sebesar 0,69 masih tergolong moderat, menunjukkan tantangan dalam mendeteksi keluhan pengguna yang jumlahnya relatif sedikit.

Kegagalan dalam mendeteksi sentimen netral sebagian besar disebabkan oleh ketimpangan distribusi label dalam dataset. Label netral hanya mencakup 26 dari 2.000 ulasan, atau sekitar 1,3%, sehingga jumlah datanya tidak cukup untuk membentuk pola klasifikasi yang kuat. Dalam kondisi seperti ini, algoritma SVM yang secara default tidak mempertimbangkan bobot kelas akan cenderung mengabaikan kelas minoritas. Ketimpangan distribusi label ini merupakan kendala utama yang juga dilaporkan oleh Iskandar & Nataliani [12] dalam studi perbandingan SVM dan Naïve Bayes pada ulasan produk digital. Menurut Al Fahreza et al. [17], distribusi data yang tidak merata sangat memengaruhi pembentukan pola klasifikasi, karena algoritma cenderung menyesuaikan diri dengan label yang mendominasi. Hal serupa diungkapkan oleh Mardianto et al. [14], di mana ketidakseimbangan data membuat model gagal mendeteksi sentimen bunuh diri pada platform X.

Berbagai penelitian menyarankan teknik penanganan imbalanced dataset sebagai langkah solusi. Alfandi Safira & Hasan [9] menerapkan undersampling dan SMOTE untuk mengoptimalkan akurasi klasifikasi Paylater. Supian et al. [10] menunjukkan kombinasi SVM dengan Word2Vec dan SMOTE meningkatkan prediksi ulasan minoritas pada aplikasi e-wallet DANA. Sejalan dengan itu, Husada & Paramita [15] menegaskan bahwa metode penyeimbangan data dapat meningkatkan recall kelas minoritas pada analisis sentimen maskapai penerbangan.

Selain masalah distribusi, pendekatan automatic labeling menggunakan rating bintang juga memiliki kelebihan dan kekurangan. Di satu sisi, cara ini praktis dan cepat, seperti dibuktikan pada penelitian Sejati et al. [18] yang merekomendasikan automatic labeling untuk ulasan di media sosial. Namun di sisi lain, kualitas anotasi tidak selalu akurat. Beberapa pengguna bisa memberikan rating 3 tetapi dengan komentar negatif, atau sebaliknya. Hal ini sesuai dengan catatan Singgalen [19], yang menjelaskan perlunya manual check untuk validasi sampel.

Untuk meminimalkan bias, Pratiwi et al. [11] pada klasifikasi kesehatan mental remaja menambahkan manual labeling pada sebagian data untuk memvalidasi pelabelan otomatis. Alternatif lain, menurut Alita & Rahman [20], adalah menambahkan aspek sentiment score berbasis leksikon agar label lebih mewakili opini sebenarnya.

Dalam konteks metode, penerapan kernel linear pada SVM terbukti relevan untuk data teks pendek seperti ulasan aplikasi [2], [12], [13]. Namun, Mardianto et al. [14] juga mengingatkan bahwa performa SVM bisa kalah dari algoritma lain seperti Random Forest atau Logistic Regression bila data sangat tidak seimbang. Kavabilla et al. [2] bahkan membandingkan SVM dengan Maximum Entropy dan mendapati hasil serupa untuk teks ulasan keuangan.

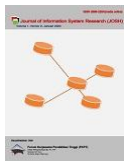
Hasil penelitian ini juga mengonfirmasi perlunya validasi silang (cross validation) yang proporsional. Supian et al. [10] menggunakan Stratified K-Fold yang serupa, terbukti mampu menjaga proporsi label di setiap fold. Tanpa stratifikasi, distribusi kelas minoritas bisa semakin hilang pada data latih, sehingga prediksi makin bias.

Dari hasil pengujian, dapat dilihat bahwa model gagal mendeteksi sentimen netral. Hal ini konsisten dengan temuan Nufairi et al. [6] pada analisis sentimen aplikasi Threads, di mana label netral hampir tidak terdeteksi akibat jumlahnya yang terlalu sedikit. Singgalen [19] juga menyoroti tantangan serupa pada data pariwisata Indonesia yang mayoritas sentimen positif.

Selain itu, hasil evaluasi ini sekaligus memperlihatkan bahwa pipeline pra-pemrosesan, transformasi TF-IDF, dan strategi validasi silang yang diterapkan sudah sesuai dengan praktik umum dalam analisis sentimen berbasis teks. Meski demikian, ketidakseimbangan data tetap menjadi tantangan mendasar yang membuka peluang riset lanjutan untuk mengintegrasikan metode balancing data secara otomatis pada tahap pelatihan model.

Solusi jangka panjang yang disarankan adalah mengombinasikan pendekatan SVM dengan deep learning atau model berbasis transformer seperti BERT [21]. Menurut Nurian & Sari [22], BERT lebih sensitif dalam memahami konteks kalimat dan efektif mendeteksi opini netral yang tidak ekstrem. Saputra & Hasan [11] juga merekomendasikan aspect-based sentiment analysis untuk menggali aspek spesifik dari layanan, misalnya kualitas konseling, kecepatan respon, atau tampilan antarmuka.

Dengan demikian, penelitian ini menegaskan bahwa penerapan SVM cocok untuk klasifikasi dua kelas dominan, tetapi tidak ideal untuk multi-kelas dengan kelas minoritas yang sangat kecil. Oleh karena itu, penelitian



mendatang perlu mengeksplorasi teknik penyeimbangan data, penyesuaian bobot kelas (class weight balancing), serta ensemble learning seperti yang diuraikan oleh Wibowo et al. [23].

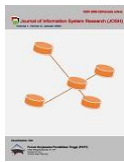
Secara keseluruhan, penelitian ini diharapkan menjadi acuan bagi pengembang layanan mental health app lokal agar dapat mendengar suara pengguna secara objektif, cepat, dan terukur. Dengan analisis sentimen yang akurat, pengembang dapat merancang perbaikan layanan yang tepat sasaran, meningkatkan kualitas fitur, dan menjaga kepercayaan pengguna di tengah meningkatnya kebutuhan layanan kesehatan mental digital di Indonesia.

4. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Riliv: Mental Health App di Google Play Store dengan menerapkan algoritma Support Vector Machine (SVM). Seluruh tahapan penelitian, mulai dari pengambilan data melalui web scraping, pra-pemrosesan teks, pelabelan otomatis berdasarkan rating, transformasi TF-IDF, pembagian data dengan Stratified K-Fold Cross Validation, hingga pelatihan dan evaluasi model, telah dilaksanakan secara sistematis. Hasil yang diperoleh menunjukkan bahwa model SVM cukup efektif dalam mengklasifikasikan sentimen positif dan negatif dengan akurasi keseluruhan mencapai 85,56% dan nilai f1-score tertinggi pada label positif sebesar 0,93. Namun, penelitian ini memiliki keterbatasan signifikan pada klasifikasi sentimen netral karena distribusi data yang sangat tidak seimbang, di mana hanya sebagian kecil ulasan berlabel netral. Hal ini membuat model gagal mengenali opini netral dengan tepat. Secara keseluruhan, model SVM efektif dalam menangani dua kelas utama sentimen yang umum dijumpai, yaitu positif dan negatif. Namun, untuk klasifikasi multikelas dengan keberadaan kelas minoritas, dibutuhkan pendekatan tambahan guna meningkatkan kemampuan model dalam menangani distribusi yang tidak merata. Oleh karena itu, penelitian selanjutnya disarankan untuk menerapkan teknik penyeimbangan data seperti oversampling, undersampling, atau Synthetic Minority Oversampling Technique (SMOTE), serta mempertimbangkan penggunaan parameter `class_weight` pada SVM. Selain itu, eksplorasi algoritma alternatif seperti Random Forest atau model berbasis transformer seperti BERT diharapkan mampu meningkatkan akurasi pada data multikelas yang tidak seimbang. Pengembangan ke depan juga dapat diarahkan pada penerapan aspect-based sentiment analysis agar dapat menggali persepsi pengguna secara lebih mendalam dan kontekstual, sehingga hasil analisis dapat menjadi masukan yang lebih komprehensif bagi pengembangan layanan digital kesehatan mental.

REFERENCES

- [1] S. Gunardi et al., Pengantar Sistem Informasi. MEGA PRESS NUSANTARA, 2024.
- [2] F. E. Kavabilla, T. Widiyari, dan B. Warsito, "Analisis Sentimen Pada Ulasan Aplikasi Investasi Online Ajaib Pada Google Play Menggunakan Metode Support Vector Machine Dan Maximum Entropy," *J. Gaussian*, vol. 11, no. 4, hal. 542–553, Feb 2023, doi: 10.14710/j.gauss.11.4.542-553.
- [3] R. Saputra dan F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, hal. 411–419, 2024.
- [4] F. N. Hasan dan M. Dwijayanti, "Analisis Sentimen Ulasan Pelanggan Terhadap Layanan Grab Indonesia Menggunakan Multinomial Naive Bayes Classifier," *J. Linguist. Komputasional*, vol. 4, no. 2, 2021.
- [5] A. B. Muzayyanah, R. E. Pawening, dan Z. Arifin, "Analisis Sentimen Pada Ulasan Aplikasi Ehadrah Di Google Playstore Menggunakan Support Vector Machine (Svm)," *Idealis Indones. J. Inf. Syst.*, vol. 7, no. 2, hal. 258–266, Jul 2024, doi: 10.36080/idealis.v7i2.3250.
- [6] F. Nufairi, N. Pratiwi, dan F. Herlando, "Analisis Sentimen Pada Ulasan Aplikasi Threads Di Google Play Store Menggunakan Algoritma Support Vector Machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, hal. 339–348, Feb 2024, doi: 10.29100/jipi.v9i1.4929.
- [7] E. Eskiyyaturofikoh dan R. R. Suryono, "Analisis Sentimen Aplikasi X Pada Google Play Store Menggunakan Algoritma Naive Bayes Dan Support Vector Machine (SVM)," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, hal. 1408–1419, Agu 2024, doi: 10.29100/jipi.v9i3.5392.
- [8] E. E. Pratiwi, A. R. Aisy, R. Rahmaddeni, dan N. Ananta, "Klasifikasi Kesehatan Mental Pada Usia Remaja Menggunakan Metode SVM," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 2, Apr 2025, doi: 10.23960/jitet.v13i2.6232.
- [9] Alfandi Safira dan F. N. Hasan, "Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier," *Zo. J. Sist. Inf.*, vol. 5, no. 1, hal. 59–70, Jan 2023, doi: 10.31849/zn.v5i1.12856.
- [10] A. Supian, B. T. Revaldo, N. Marhadi, dan L. Efrizoni, "Penerapan SVM dan Word2Vec untuk Analisis Sentimen Ulasan Pengguna Aplikasi DANA," *J. Ilm. Komputasi*, vol. 23, no. 3, hal. 321–332, Sep 2024, doi: 10.32409/jikstik.23.3.3642.
- [11] A. Saputra dan F. Noor Hasan, "Analisis Sentimen Terhadap Aplikasi Coffee Meets Bagel Dengan Algoritma Naive Bayes Classifier," *SIBATIK J. J. Ilm. Bid. Sos. Ekon. Budaya, Teknol. dan Pendidik.*, vol. 2, no. 2, hal. 465–474, Jan 2023, doi: 10.54443/sibatik.v2i2.579.
- [12] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naive Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 6, hal. 1120–1126, Des 2021, doi: 10.29207/resti.v5i6.3588.
- [13] O. I. Gifari, M. Adha, F. Freddy, dan F. F. S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, hal. 36–40, Mar 2022, doi: 10.46229/jifotech.v2i1.330.
- [14] M. F. F. Mardianto, B. S. Pratama, M. Audilla, dan E. Pusporani, "Sentiment Analysis of Suicide on X Using Support Vector Machine and Naive Bayes Classifier Algorithms," *INTENSIF J. Ilm. Penelit. dan Penerapan Teknol. Sist. Inf.*,



- vol. 9, no. 1, hal. 60–75, Feb 2025, doi: 10.29407/intensif.v9i1.23742.
- [15] H. C. Husada dan A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, hal. 18–26, Feb 2021, doi: 10.34148/teknika.v10i1.311.
- [16] Y. A. Singgalen, “Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia,” *J. Inf. Syst. Informatics*, vol. 3, no. 3, hal. 459–470, Okt 2021, doi: 10.51519/journalisi.v3i3.171.
- [17] Muhammad Daffa Al Fahreza, Ardytha Luthfiarta, Muhammad Rafid, dan Michael Indrawan, “Analisis Sentimen: Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z,” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, hal. 16–25, Feb 2024, doi: 10.52158/jacost.v5i1.715.
- [18] Wahyu Sejati, Ankur Singh Bist, dan Amirsyah Tambunan, “Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis,” *J. MENTARI Manajemen, Pendidik. dan Teknol. Inf.*, vol. 2, no. 1, hal. 95–103, Agu 2023, doi: 10.33050/mentari.v2i1.377.
- [19] Y. A. Singgalen, “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Sistic Literature Review,” *J. Inf. Syst. Informatics*, vol. 3, no. 2, hal. 278–302, Jun 2021, doi: 10.33557/journalisi.v3i2.125.
- [20] D. Alita dan A. R. Isnain, “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier,” *J. komputasi*, vol. 8, no. 2, hal. 50–58, Okt 2020, doi: 10.23960/komputasi.v8i2.2615.
- [21] Septi Putri, Yohanes Agung Apriyanto, dan Andri Wijaya, “Analisis Sentimen Ulasan Aplikasi Deepl Pada Google Play Dengan Metode Support Vector Machine (SVM),” *J. Sist. Inf.*, vol. 4, no. 2, hal. 59–66, Des 2023, doi: 10.32546/jusin.v4i2.2368.
- [22] A. Nurian, “Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3s1, Sep 2023, doi: 10.23960/jitet.v11i3s1.3348.
- [23] A. Wibowo, Firman Noor Hasan, Rika Nurhayati, dan Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asimetrik J. Ilm. Rekayasa Inov.*, hal. 239–248, Jul 2022, doi: 10.35814/asimetrik.v4i1.3577.