Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

EKSPLORASI PRA-PROSES LARGE LANGUAGE MODEL GEMINI 2.0 UNTUK KLASIFIKASI ULASAN SIREKAP MENGGUNAKAN SVM

Muhamad Ridwan[1], Naufal Azmi Verdikha[2], Muhammad Albert Putra Firman[3]
1, 2 ,3) Teknik Informatika, Fakultas Fakultas Sains dan Teknologi,
Universitas Muhammadiyah Kalimantan Timur, Indonesia

Article Info

Article history:

Received: 30 Juli 2026
Revised: 01 Agustus 2026
Accepted: 03 Agustus 2026

ABSTRACT

Abstrak

Implementasi aplikasi SIREKAP 2024 dalam Pemilihan Umum menuai beragam respon publik. Ulasan Google Play Store menjadi data krusial untuk mengevaluasi kepuasan pengguna melalui rating. Studi ini bertujuan mengklasifikasikan ulasan SIREKAP ke dalam lima kategori rating. Metode yang digunakan adalah Support Vector Machine (SVM) dengan pembobotan TF-IDF. Kebaruan penelitian ini terletak pada pemanfaatan API Gemini 2.0 Flash (LLM) untuk tahap pra-proses, yang menangani koreksi ejaan dan normalisasi kata secara kontekstual. Analisis menunjukkan ketidakseimbangan kelas yang signifikan, didominasi oleh rating 1 sebesar 64,24%. Berdasarkan pengujian K-Fold Cross Validation, model LinearSVC menghasilkan rata-rata F1-Score sebesar 36,68%. Skor ini menunjukkan kesulitan model dalam memprediksi kelas, sehingga memerlukan teknik penanganan lanjutan dalam penelitian mendatang untuk meningkatkan performa..

Kata Kunci: Gemini, Large Language Model (LLM), SIREKAP 2024, Support Vector Machine (SVM).

Abstract

The implementation of the 2024 SIREKAP application in the General Election garnered various public responses. Google Play Store reviews serve as crucial data for evaluating user satisfaction via assigned ratings. Consequently, this study aims to classify SIREKAP reviews into five rating categories. The method employed is Support Vector Machine (SVM) with TF-IDF weighting. A key novelty is utilizing the Gemini 2.0 Flash API (LLM) for preprocessing, effectively handling spelling correction and contextual word normalization. Analysis reveals a significant class imbalance, heavily dominated by rating 1 at 64.24%. Using K-Fold Cross Validation, the LinearSVC model yielded an average F1-Score of 36.68%. This score indicates the model struggled with class prediction, highlighting the necessity for advanced handling techniques in future research to improve performance.

Keywords: Gemini, Large Language Model (LLM), SIREKAP 2024, Support Vector Machine (SVM).

Corresponding Author:
E-mail : nav651@umkt.ac.id

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

1. PENDAHULUAN

Aplikasi SIREKAP 2024 merupakan instrumen krusial berbasis teknologi informasi untuk publikasi hasil pemilu yang transparan, di mana ulasan pengguna pada Google Play Store menjadi indikator penting dalam penelitian ini [1], [2]. Namun, analisis ulasan ini menghadapi tantangan besar berupa data tidak terstruktur yang didominasi bahasa tidak baku dan kesalahan pengetikan ekstrem, yang seringkali menyebabkan ketidaksesuaian antara narasi teks dengan rating yang diberikan [3]. Masalah kualitas data ini terbukti menghambat performa klasifikasi, sebagaimana ditunjukkan oleh penelitian sebelumnya yang mengombinasikan DistilBERT dan Support Vector Machine (SVM) namun hanya menghasilkan F1-score 36,62% [4]. Hasil ini sebagian besar disebabkan oleh metode pembersihan manual atau berbasis aturan kaku yang justru menghilangkan informasi penting saat normalisasi dilakukan [5].

Untuk mengatasi keterbatasan metode pra-proses konvensional yang hanya bergantung pada kamus statis dan gagal menangkap makna teks penuh gangguan semantik [6], penelitian ini menerapkan teknologi Large Language Models (LLM) melalui model Gemini. Pendekatan ini dipilih karena kemampuan generatif LLM dalam memahami konteks kalimat memungkinkan koreksi kesalahan ejaan dan tata bahasa yang jauh lebih akurat dibandingkan algoritma konvensional [7], . Intervensi LLM Gemini dalam tahap pra-proses menjadi strategi kebaruan utama penelitian ini untuk menjamin kualitas data teks sebelum dikonversi menjadi representasi numerik, sejalan dengan bukti bahwa perbaikan anomali teks berbasis AI efektif meningkatkan kualitas dataset [8].

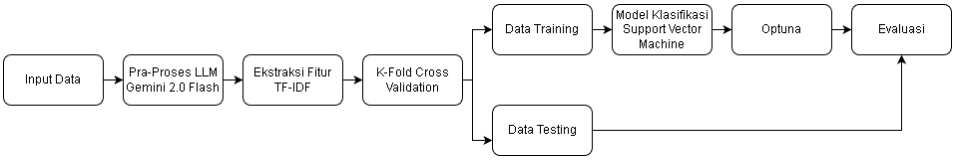
Setelah kualitas data ditingkatkan, proses klasifikasi memerlukan konversi teks menjadi format vektor numerik agar dapat diproses secara matematis [9], [10]. Penelitian ini menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk ekstraksi fitur, mengingat efektivitasnya yang terbukti mampu menghasilkan akurasi tinggi hingga 96,63% dalam studi terdahulu [11], [12]. Selanjutnya, klasifikasi dilakukan menggunakan algoritma Support Vector Machine (SVM). Pemilihan SVM didasarkan pada keunggulannya dalam menangani data teks berdimensi tinggi, di mana studi komparasi menunjukkan bahwa SVM secara konsisten mengungguli metode lain seperti Naïve Bayes dengan selisih akurasi yang signifikan [13], [14].

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Secara keseluruhan, penelitian ini bertujuan untuk menguji efektivitas kombinasi pra-proses berbasis LLM Gemini, ekstraksi fitur TF-IDF, dan klasifikasi SVM terhadap ulasan aplikasi SIREKAP 2024 [15]. Mengingat karakteristik ketidakseimbangan kelas pada data ulasan, evaluasi model difokuskan pada metrik F1-Score. Metrik ini dipilih karena kemampuannya menggambarkan dataset dengan karakteristik kelas yang tidak seimbang, sehingga memberikan evaluasi performa yang lebih komprehensif dan objektif terhadap kemampuan model dalam memprediksi setiap kategori rating [16].

2. METODE PENELITIAN

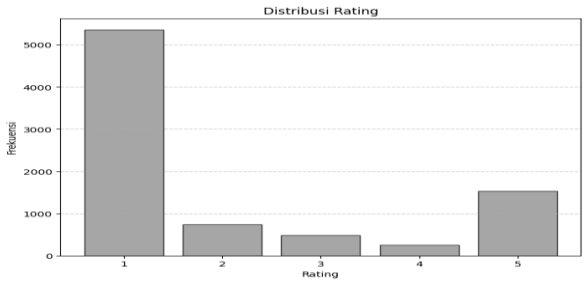
Penelitian ini dirancang menggunakan kerangka kerja metodologis yang terstruktur untuk menjamin validitas hasil klasifikasi ulasan. Alur penelitian secara sistematis dimulai dari tahap pengumpulan data, dilanjutkan dengan normalisasi teks, transformasi fitur, hingga tahap pengujian performa model. Seluruh rangkaian prosedur yang diterapkan dalam penelitian ini diilustrasikan secara mendalam pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Input Data

Tahapan penelitian ini diawali dengan proses input data yang menggunakan data sekunder yang bersumber dari penelitian sebelumnya yang terdapat 8.358 ulasan pengguna aplikasi SIREKAP 2024 diambil dari Google Play Store menggunakan metode Scraping pada tanggal 6 Februari 2024 [4]. Data tersebut dikumpulkan melalui metode scraping pada laman Google Play Store dan tersimpan dalam format berkas Comma-Separated Values (CSV). Distribusi data mayoritas ulasan pengguna memberikan penilaian seperti pada Gambar 2.



Gambar 1. Distribusi Dataset

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Seperti terlihat pada Tabel 2, dominasi data pada kelas Rating 1 mencapai 64,24%, sedangkan Rating 4 hanya memiliki proporsi 3,1%. Ketidakseimbangan ini menjadi tantangan utama dalam proses klasifikasi yang harus ditangani melalui pemilihan metode evaluasi yang tepat.Pada umumnya jurnal tidak menginginkan Bahasa statistic (seperti: *significantly different*, *treatment*, dll) ditulis dalam pembahasan. Hindari*copy*dan *paste*tabel hasil analisis statistik langsung dari *software*pengolah data statistik.

3.2 Arsitektur Pra-proses

Model Gemini 2.0 Flash dipilih karena efisiensi komputasinya dalam menangani tugas pra-proses teks [17]. Tahap ini dirancang untuk mentransformasi ulasan mentah menjadi data terstruktur yang bersih menggunakan empat komponen arsitektur utama: inisialisasi API, manajemen penyimpanan sementara, penerapan protokol instruksi, dan eksekusi tahapan sistematis. Fokus utama arsitektur ini adalah menjamin efisiensi penggunaan kuota, konsistensi format, dan validitas hasil pemrosesan teks.

Inisialisasi API Gemini 2.0 Flash dilakukan menggunakan pustaka Python google-generativeai versi 0.8.5 sebagai jembatan komunikasi dengan layanan Generative AI milik Google. Pemilihan model Gemini 2.0 Flash didasarkan pada efisiensi komputasi, namun memerlukan penanganan khusus terhadap batasan sumber daya pada kapasitas free tier. Batasan 15 Requests Per Minute (RPM) menjadi kendala operasional utama yang diatasi dengan merancang algoritma pengiriman data menggunakan mekanisme time delay otomatis guna menjaga stabilitas sistem di bawah ambang batas server. Rincian spesifikasi batasan kuota API dirinci pada Tabel 2.

Tabel 1. Spesifikasi Batasan Kuota API Gemini 2.0 Free Tier

Jenis Ukuran	Limitasi Maksimal	Penjelasan
Rate Limit (RPM)	15 Requests Per Minute	Batas kecepatan pengiriman permintaan ke server Google, yaitu maksimal 15 kali dalam satu menit.
Daily Limit (RPD)	1.500 Requests Per Day	Batas total jumlah permintaan yang dapat diproses secara gratis dalam kurun waktu 24 jam.
Token Limit (TPM)	1.000.000 Tokens Per Minute	Batas jumlah unit teks (token) yang dapat diproses dalam satu menit, mencakup gabungan teks input dan output.

Untuk mengatasi tantangan latensi jaringan dan menjaga progres pengerjaan agar tidak hilang saat program dihentikan diterapkan manajemen cache berbabis JavaScript Object Notation (JSON). Strategi penyimpanan lokal sementara berbasis berkas ini

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

memastikan setiap data yang berhasil melalui tahapan pra-proses tersimpan secara permanen dalam format JSON. Alur logika sistem mencakup: (1) pengecekan ketersediaan berkas di folder cacheI lokal, (2) pemuatan data langsung menggunakan fungsi json.load() jika data tersedia guna Menghemat kuota harian, dan (3) penyimpanan respons baru dari API menggunakan fungsi json.dump() untuk kebutuhan tahapan selanjutnya.

Prompt Protokol atau Mandatory Rule Prompts digunakan untuk mengendalikan model agar bertindak sebagai mesin pemroses data presisi dan menekan sifat generatif alami model (halusinasi). Protokol ini disisipkan pada setiap batch permintaan untuk memaksa model mematuhi aturan ketat: (1) output wajib dalam format JSON Dictionary murni tanpa narasi tambahan, (2) jumlah item keluaran harus sama persis dengan jumlah input guna mencegah pergeseran indeks data asli, serta (3) penggunaan instruksi sistem untuk menetapkan peran spesifik model pada setiap tugas pembersihan.

Tabel 2. Komponen Input dan Output Token dalam Satu Request

Kategori	Komponen	Keterangan
Input Token	System Role	Definisi peran model (Contoh: "Anda adalah mesin pemroses teks...")
	Task Instruction	Perintah spesifik tahap terkait (Contoh: "Ubah ke huruf kecil...")
	Output Rules	Aturan format JSON dan larangan teks narasi
Output Token	*JSON Response*	Hasil olahan dari 50 data tersebut dalam format JSON murni

2.3 Tahapan Pra-proses

Tahapan pra-proses dilakukan secara berurutan untuk mentransformasi teks ulasan mentah menjadi korpus yang berkualitas tinggi sebelum tahap ekstraksi fitur. Proses ini melibatkan empat langkah utama yang dijalankan melalui instruksi prompt pada model Gemini pada Tabel 3.

Tabel 3. Prosedur Tahapan Pra-proses Data

Tahap	Deskripsi Singkat	Contoh transformasi
Lowercase	Mengubah seluruh karakter menjadi huruf kecil untuk meniadakan perbedaan makna akibat kapitalisasi.	“Mantap Sekali” -> “mantap sekali”
Remove Character Unnecessary	Menghapus angka, simbol, tanda baca, dan emoji yang tidak memiliki nilai informasi klasifikasi.	“:error!!! 123 😡” -> “error”

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Spellchecker	Memperbaiki kesalahan ketik (typo) serta menormalisasi singkatan dan bahasa gaul menjadi kata baku sesuai konteks kalimat Bahasa Indonesia.	"apk nggk bisa log in" -> "aplikasi tidak bisa masuk"
Stemming	Mereduksi kata berimbuhan menjadi kata dasar murni untuk meminimalisir dimensi fitur tanpa menghilangkan makna dasar ulasan.	"membantu" -> "bantu"

2.4 Ekstraksi Fitur TF-IDF

Data teks yang telah dibersihkan dikonversi menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini dipilih karena kemampuannya dalam memberikan bobot tinggi pada kata-kata penting yang unik dalam suatu ulasan. Implementasi ekstraksi fitur dilakukan menggunakan fungsi TfidfVectorizer dari pustaka Scikit-learn dengan konfigurasi parameter sebagai berikut: (a) ngram_range diatur pada (1, 1) agar sistem hanya mengambil kata tunggal sebagai fitur, (b) parameter norm menggunakan nilai 'l2' untuk memastikan seluruh vektor memiliki panjang yang seragam, dan (c) parameter use_idf diatur pada nilai True untuk mengaktifkan pembobotan frekuensi dokumen terbalik.

2.5 Klasifikasi dan Optimasi SVM

Algoritma Support Vector Machine (SVM) digunakan sebagai model utama untuk mengklasifikasikan ulasan ke dalam kategori rating. Model ini diimplementasikan melalui fungsi LinearSVC dengan kernel linear karena karakteristik data hasil ekstraksi TF-IDF yang memiliki dimensi tinggi cenderung dapat dipisahkan secara linier. Guna mencapai performa maksimal, dilakukan prosedur hyperparameter tuning menggunakan kerangka kerja Optuna melalui 200 trials. Parameter kunci yang dioptimasi dalam proses ini mencakup nilai regularisasi (C), batas maksimum iterasi (max_iter), serta pengaturan bobot kelas (class_weight) dengan nilai balanced untuk menangani ketidakseimbangan distribusi data.

2.6 Metrik Evaluasi

Kinerja model divalidasi menggunakan skema 10-Fold Cross Validation untuk menjamin stabilitas hasil evaluasi dan menghindari kondisi overfitting [19]. Dalam teknik ini, dataset dibagi menjadi sepuluh subset berukuran sama yang digunakan secara bergantian sebagai data latih dan data uji. Metrik evaluasi utama yang dipilih adalah F1-

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Score. Pemilihan metrik ini didasarkan pada efektivitasnya dalam mengukur keseimbangan antara precision dan recall.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Hasil Pra-proses Data

Implementasi pra-proses menggunakan Gemini 2.0 Flash berhasil mentransformasi ulasan mentah menjadi korpus teks yang bersih dan terstandardisasi. Dari total 8.358 data ulasan awal, terdapat 44 data yang dieliminasi karena hanya berisi karakter non-alfabet (seperti emoji murni) yang hilang setelah pembersihan, sehingga menyisakan 8.314 data ulasan. Meskipun terjadi penyusutan jumlah baris, karakteristik distribusi rating tetap menunjukkan ketimpangan yang signifikan, di mana Rating 1 mendominasi dataset sebesar 64,24%. Rincian distribusi ulasan disajikan pada Tabel 4.

Tabel 4. Distribusi Rating Sebelum dan Sesudah Pra-proses

Rating	Sebelum pra-proses	Sesudah pra-proses	Persentase
1	5.361	5.341	64,24%
2	736	736	8,85%
3	478	478	5,75%
4	255	254	3,06%
5	1.528	1.505	18,10%
Total	8.358	8.314	100%

Secara kualitatif, integrasi LLM terbukti efektif dalam menangani anomali teks yang kompleks. Sebagai contoh, ulasan "apk nggk bisa buat log in" berhasil dinormalisasi menjadi "aplikasi tidak bisa buat masuk" melalui tahap spellchecker. Hal ini menunjukkan keunggulan pendekatan berbasis LLM dalam memahami konteks semantik kalimat dibandingkan metode berbasis aturan tradisional.

3.2 Hasil Optimasi Parameter

Optimasi hiperparameter mengotomatiskan penyesuaian nilai hiperparameter untuk membuat proses lebih efisien. Secara keseluruhan, tujuan optimasi hyperparameter adalah untuk mencapai efisiensi ini [20]. Optimasi hyperparameter menggunakan Optuna dilakukan untuk menemukan konfigurasi LinearSVC terbaik [18]. Berdasarkan 200 trials, model mencapai titik performa optimal pada trial ke-122 dengan nilai parameter regularisasi C=0,1618. Nilai C yang relatif kecil ini mengindikasikan

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

bahwa model memerlukan margin yang lebih lebar untuk menoleransi gangguan (noise) pada data ulasan guna mencegah terjadinya overfitting

3.3 Performa Model Klasifikasi

Evaluasi kinerja model menggunakan 10-Fold Cross Validation menghasilkan rata-rata Macro F1-Score sebesar 36,68%. Nilai ini mencerminkan kemampuan generalisasi model terhadap seluruh subset data uji yang digunakan. Rincian hasil pengukuran metrik untuk setiap iterasi disajikan pada Tabel 5.

Tabel 5. Hasil Evaluasi F1-Score Tiap Fold

Tabel 5. Hasil Evaluasi F1-Score Tiap Fold

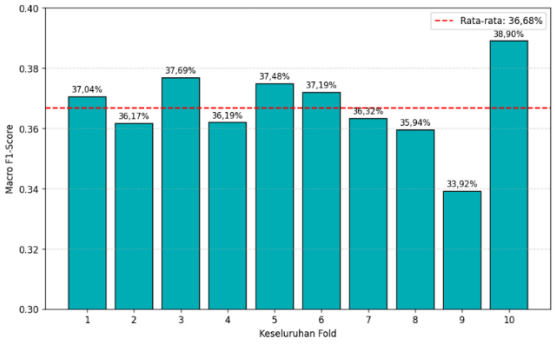
Fold	F1_Score
1	0,3704
2	0,3617
3	0,3769
4	0,3619
5	0,3748
6	0,3719
7	0,3632
8	0,3594
9	0,3392
10	0,3890
Rata-rata	0,3668

Berdasarkan Tabel 5 di atas, rata-rata nilai F1-Score yang diperoleh dari keseluruhan fold adalah 36,68%. Nilai ini dihitung menggunakan metode Macro Average, yang merata-ratakan performa kelima kategori rating secara setara tanpa memandang proporsi jumlah data. Sebagai konsekuensinya, rendahnya akurasi pada kelas rating 2, 3, dan 4 secara langsung menarik turun nilai rata-rata keseluruhan, meskipun model terbukti mampu memprediksi kelas rating 1 dengan baik.

Fenomena rendahnya performa pada kelas-kelas tersebut disebabkan oleh tingginya ambiguitas semantik pada data ulasan. Berdasarkan analisis data, ditemukan banyak keluhan identik yang tersebar secara tidak konsisten di rating yang berbeda. Contohnya, frasa "susah masuk" muncul pada ulasan rating 1 dan rating 2, sedangkan keluhan "gagal masuk terus" justru muncul pada rating 3. Tumpang tindih (overlapping) fitur linguistik ini menyulitkan algoritma SVM untuk menarik garis pemisah (hyperplane) yang tegas antar kelas.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

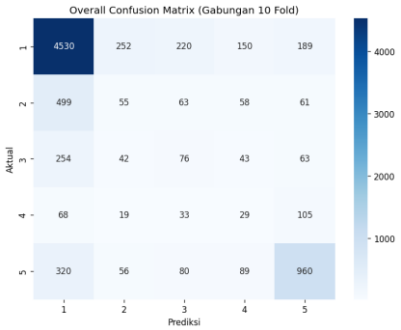
Untuk memberikan gambaran visual mengenai stabilitas performa model, Gambar 3 menyajikan grafik F1-Score dari keseluruhan fold. Grafik ini memperlihatkan fluktuasi kinerja model pada setiap subset data uji, serta garis rata-rata yang merepresentasikan kemampuan generalisasi model secara keseluruhan.



Gambar 3. Grafik Rata-rata F1-Score Keseluruhan Fold

Stabilitas performa model yang tervisualisasi pada grafik di atas menunjukkan bahwa meskipun terdapat fluktuasi nilai F1-Score antar fold, rentang perbedaannya relatif terkendali, yakni berkisar antara 33,92% hingga 38,90%. Pola distribusi ini mengonfirmasi bahwa model memiliki tingkat generalisasi yang cukup stabil terhadap variasi pembagian data, namun juga secara konsisten menghadapi kendala yang seragam dalam memisahkan kelas rating pada setiap iterasinya. Stagnasi performa pada angka rata-rata 36,68% ini bukanlah anomali pada satu subset data tertentu, melainkan cerminan dari tantangan struktural yang dihadapi model terhadap karakteristik data secara keseluruhan.

Guna mengevaluasi detail kesalahan prediksi secara komprehensif, hasil klasifikasi dari seluruh 10-fold digabungkan ke dalam sebuah Overall Confusion Matrix. Visualisasi ini memetakan perbandingan antara label rating aktual (sumbu Y) dengan label rating yang diprediksi oleh model (sumbu X) untuk mengidentifikasi kelas mana yang paling dominan mengalami kesalahan klasifikasi.



Gambar 4. Overall Confusion Matrix Klasifikasi Ulasan SIREKAP 2024

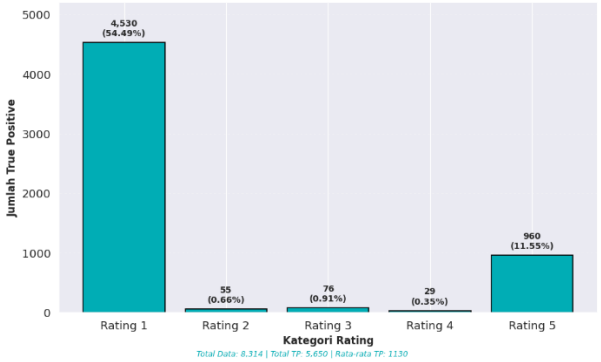
Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Berdasarkan visualisasi pada Gambar 4, model menunjukkan kemampuan yang sangat tinggi dalam mengenali ulasan dengan rating 1, yang merupakan kelas mayoritas dalam dataset. Hal ini dibuktikan dengan jumlah True Positive mencapai 4.530 data. Sebaliknya, performa model menurun drastis pada kelas minoritas, dengan jumlah prediksi benar terendah ditemukan pada rating 4 yang hanya berjumlah 29 data. Rincian distribusi prediksi benar untuk setiap kategori rating dirangkum dalam Tabel 6.

Tabel 6. Rincian Hasil Prediksi Benar (True Positive)

Kategori Rating	Prediksi Benar (True Positive)
Rating 1	4.530
Rating 2	55
Rating 3	76
Rating 4	29
Rating 5	960

Selain prediksi benar, analisis terhadap misclassification menunjukkan adanya pola kesalahan yang menonjol pada kategori rating menengah. Kesalahan paling signifikan terjadi pada kelas rating 2, di mana sebanyak 449 data salah diprediksi sebagai rating 1. Pola serupa ditemukan pada rating 5, dengan 320 data yang keliru diklasifikasikan ke dalam rating 1. Kecenderungan model untuk memprediksi ulasan ke dalam kelas rating 1 disebabkan oleh dominasi fitur-fitur linguistik dari kelas mayoritas tersebut yang mengaburkan hyperplane pada algoritma SVM.



Gambar 5. Distribusi Prediksi Benar (True Positive)

Gambar 5 memvisualisasikan ketimpangan distribusi prediksi benar antar kelas. Terlihat jelas bahwa batang untuk Rating 1 (4.530 data) mendominasi visualisasi, sementara batang untuk Rating 2, 3, dan 4 sangat pendek. Fenomena ini secara kuantitatif menunjukkan bahwa model hampir selalu memprediksi ulasan ke dalam kelas mayoritas.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Perbedaan antara Rating 1 dan Rating 4 mencapai 4.501 data, yang mengindikasikan ketimpangan performa yang ekstrem. Jika dibandingkan dengan rata-rata prediksi benar per kelas (1.130), hanya Rating 1 dan Rating 5 yang berada di atas rata-rata, sementara tiga kelas lainnya (Rating 2, 3, dan 4) berada jauh di bawah. Hal ini mengonfirmasi bahwa model gagal dalam menangani kelas minoritas, yang merupakan masalah klasik pada dataset dengan ketidakseimbangan kelas yang parah.

Persentase prediksi benar terhadap total data juga menunjukkan ketimpangan Rating 1 mencapai 54,5% dari total data (4.530/8.314), sementara Rating 4 hanya 0,3% (29/8.314). Ini berarti hampir seluruh kemampuan klasifikasi model terserap oleh kelas Rating 1, meninggalkan kelas-kelas lainnya dengan performa yang sangat buruk.

4. SIMPULAN

Berdasarkan hasil analisis dan tahapan eksperimen yang telah dilakukan, penelitian ini menghasilkan beberapa poin simpulan dan arahan pengembangan sebagai berikut. (1) Integrasi model Gemini 2.0 Flash API sebagai instrumen pra-proses terbukti memberikan kontribusi signifikan dalam meningkatkan kualitas dataset ulasan SIREKAP 2024. Kemampuan LLM dalam menangani normalisasi ejaan (spellchecker) dan pembersihan teks secara kontekstual jauh lebih efektif dibandingkan metode berbasis aturan (rule-based) tradisional, terutama pada data ulasan yang didominasi bahasa tidak baku dan typo ekstrem. (2) Implementasi algoritma Linear Support Vector Machine (SVM) yang dioptimasi menggunakan Optuna mencapai titik optimal pada nilai regularisasi C = 0,1618 dengan rata-rata Macro F1-Score sebesar 36,68% melalui pengujian 10-Fold Cross Validation. (3) Hasil performa tersebut membuktikan bahwa pendekatan hibrida antara pra-proses cerdas (LLM) dan klasifikasi konvensional (SVM) mampu menghasilkan kinerja yang kompetitif dan sedikit melampaui penelitian sebelumnya yang menggunakan model Deep Learning DistilBERT (36,62%) pada dataset yang sama, namun dengan kebutuhan sumber daya komputasi yang lebih efisien. (4) Kendala utama yang ditemukan adalah adanya ketidakseimbangan kelas (class imbalance) yang sangat ekstrem, di mana Rating 1 mendominasi sebesar 64,24% dari total data, sehingga model cenderung mengalami bias dalam memprediksi kelas-kelas minoritas pada rating menengah.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno

Sebagai saran untuk penelitian di masa mendatang, terdapat beberapa langkah strategis yang dapat ditempuh guna meningkatkan akurasi klasifikasi. (5) Peneliti selanjutnya disarankan untuk menerapkan teknik penanganan ketidakseimbangan data seperti Synthetic Minority Over-sampling Technique (SMOTE) atau penyesuaian class weight yang lebih dinamis untuk memberikan bobot lebih pada kelas minoritas. (6) Pemanfaatan LLM dapat diperluas tidak hanya pada tahap pra-proses, tetapi juga untuk melakukan auto-labeling guna meminimalkan ambiguitas semantik antara teks ulasan dengan rating yang diberikan pengguna. (7) Eksplorasi penggunaan metode ekstraksi fitur yang lebih kaya seperti Word Embedding (Word2Vec atau FastText) dikombinasikan dengan arsitektur SVM diharapkan dapat menangkap makna tersembunyi dalam ulasan pengguna secara lebih presisi dibandingkan metode frekuensi kata seperti TF-IDF.

REFERENCES

[1] M. Nurkamiden, "SiRekap : Tantangan dan Potensi Kekeliruan Proses Rekapitulasi Pemilu Serentak di Indonesia SiRekap: Challenges and Potential Errors in the Recapitulation Process of Simultaneous Elections in Indonesia," 2024.

[2] I. Aida Sapitri, M. Fikry, F. Sains dan Teknologi, and U. Islam Negeri Sultan Syarif Kasim Riau, "PENGKLASIFIKASIAN SENTIMEN ULASAN APLIKASI WHATSAPP PADA GOOGLE PLAY STORE MENGGUNAKAN SUPPORT VECTOR MACHINE," Jurnal TEKINKOM, vol. 6, no. 1, 2023, doi: 10.37600/tekinkom.v6i1.773.

[3] T. Aljrees et al., "Contradiction in text review and apps rating: prediction using textual features and transfer learning," PeerJ Comput. Sci., vol. 10, 2024, doi: 10.7717/peerj-cs.1722.

[4] R. Ridhoi, N. Azmi Verdikha, and F. Yulianto, "Analisis Klasifikasi Ulasan Aplikasi Sirekap 2024 menggunakan Ekstraksi Fitur DistilBert Dan Metode Support Vector Machine," Jurnal Ilmiah Informatika (JIF), Mar. 2025.

[5] T. Bendinelli, A. Dox, and C. Holz, "Exploring LLM Agents for Cleaning Tabular Machine Learning Datasets," ICLR 2025 Workshop on Foundation Models in the Wild, Mar. 2025, [Online]. Available: http://arxiv.org/abs/2503.06664

[6] M. Vangeli, "Large Language Models as Advanced Data Preprocessors Transforming Unstructured Text into Fine-Tuning Datasets," 2024.

[7] E. Boros, M. Ehrmann, M. Romanello, S. Najem-Meyer, and F. Kaplan, "Post-correction of Historical Text Transcripts with Large Language Models: An Exploratory Study," Association for Computational Linguistics, pp. 133–159, Mar. 2024, Accessed: Oct. 29, 2025. [Online]. Available: https://aclanthology.org/2024.latechclfl-1.14/

[8] N. Schwitter, "Using large language models for preprocessing and information extraction from unstructured text: A proof-of-concept application in the social sciences," Method. Innov., vol. 18, no. 1, pp. 61–65, Mar. 2025, doi: 10.1177/20597991251313876.

[9] K. N. Singh, A. Dorendro, H. M. Devi, and A. K. Mahanta, "Analysis of Changing Trends in Textual Data Representation," in Communications in Computer and Information Science, Springer Science and Business Media Deutschland GmbH, 2021, pp. 237–251. doi: 10.1007/978-981-16-0507-9_21.

[10] M. A. Mohammed, R. A. Alsunousi, and A. Alhenshiri, "The impact of Text Representation on Classification Accuracy," African Journal of Advanced Pure and Applied Sciences, vol. 158, no. Volume 4, Issue 3, 2025, Aug. 2025, Accessed: Dec. 10, 2025. [Online]. Available: https://aaasjournals.com/index.php/ajapas/article/view/136

[11] K. Alemerien, A. Al-Ghareeb, and M. Z. Alksasbeh, "Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization," Journal of Mobile Multimedia, vol. 20, no. 5, pp. 1089–1116, 2024, doi: 10.13052/jmm1550-4646.2055.

Djtechno : Jurnal Teknologi Informasi
Vol. 7, No. 2 Agustus 2026
E-ISSN: 2745-3758, P-ISSN : 2776-8546 DOI: 10.46576/djtechno
[12] I. Apriani, Y. Sibaroni, and I. Palupi, "Perbandingan Pembobotan Fitur TF-IDF dan TF-ABS Dalam Klasifikasi Berita Online Menggunakan Support Vector Machine (SVM)," e-Proceeding of Engineering, Jun. 2023.
[13] W. Jayanata and I. Kurniawan, "Implementasi Metode Support Vector Machine (SVM) pada Model Prediksi Rating Obat Berdasarkan Ulasan Pasien," LOGIC: Jurnal Penelitian Informatika, vol. 1, no. 1, p. 14, Sep. 2023, doi: 10.25124/logic.v1i1.6410.
[14] U. Khaira, R. Aryani, and R. W. Hardian, "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Kebijakan Kemdikbudristek Mengenai Kuota Internet Selama Covid-19," Jurnal PROCESSOR, vol. 18, no. 2, Nov. 2023, doi: 10.33998/processor.2023.18.2.897.
[15] A. Pangestu, Y. T. Arifin, and R. A. Safitri, "ANALISIS SENTIMEN REVIEW PUBLIK PENGGUNA GAME ONLINE PADA PLATFORM STEAM MENGGUNAKAN ALGORITMA NAÏVE BAYES," 2023.
[16] Robert Antonius, A. R. Zulkarnain, and H. Irsyad, "Pendekatan TF-IDF, SMOTE, dan SVM dalam Klasifikasi Sentimen Masyarakat terhadap Pemblokiran Judi Online," Buletin Ilmiah Informatika Teknologi, vol. 2, no. 3, pp. 115–122, Jun. 2024, doi: 10.58369/biit.v2i3.65.
[17] A. Janakiraman, B. Ghoraani, and D. Ph, "An Empirical Comparison of Text Summarization : A Multi-Dimensional Evaluation of Large Language Models," arXiv preprint, vol. arXiv:2504, 2025.
[18] M. Ridwan and E. Utami, "Optimized Hyperparameter Tuning for Improved Hate Speech Detection with Multilayer Perceptron," Jurnal RESTI, vol. 8, no. 4, pp. 525–534, 2024, doi: 10.29207/resti.v8i4.5949.
[19] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," Jurnal Algoritma, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.
[20] L. Yang and A. Shami, "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice," Jul. 2020, doi: 10.1016/j.neucom.2020.07.061.