P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Analisis Sentimen Ulasan Aplikasi EdLink Menggunakan *Support Vector Machine* dengan SMOTE dan *Hyperparameter Tuning*

[1]Muhamad Zidan Al-Farish, [2]Setyono, [3]Hilmy Aliy Andra Putra
[1]Ilmu Komputer, Universitas Djuanda, Bogor
[2]Ilmu Komputer, Universitas Djuanda, Bogor
[3]Ilmu Komputer, Universitas Djuanda, Bogor

E-mail: [1]i.2211106@unida.ac.id, [2]setyono@unida.ac.id,[3]hilmy.aliy@unida.ac.id

ABSTRAK

Ulasan aplikasi EdLink di Google Play Store menumpuk dalam jumlah besar sehingga sulit ditinjau pengembang secara manual. Penelitian ini menganalisis sentimennya memakai Support Vector Machine yang dipadukan SMOTE dan hyperparameter tuning. Sebanyak 2.673 ulasan dikumpulkan melalui web scraping, lalu disaring menjadi 2.113 ulasan final, terdiri atas 1.275 ulasan negatif dan 838 positif. Pelabelan ditetapkan otomatis berbasis rating bintang, dan teks melewati enam tahap pra-pemrosesan termasuk penanganan kata negasi. Model diuji melalui lima skenario kumulatif dalam kerangka Stratified 10-Fold Cross Validation, dengan TF-IDF dan SMOTE dihitung ulang di tiap lipatan hanya pada data latih guna mencegah kebocoran data. Skenario terbaik menggabungkan pra-pemrosesan, SMOTE, N-gram (1,2), dan tuning, dengan Accuracy 0,8316 dan F1-Score 0,8270 memakai kernel RBF, C sebesar 10, dan Gamma 0,01. Model lebih peka pada ulasan negatif dengan Recall 0,9263 dibanding positif yang hanya 0,6874. Keluhan pengguna terpusat pada kegagalan login, kendala setelah pembaruan, dan pengiriman berkas tugas.

Kata kunci : analisis sentimen, EdLink, hyperparameter tuning, SMOTE, support vector machine

ABSTRACT

EdLink reviews on the Google Play Store accumulate in large numbers, making manual inspection by developers impractical. This study analyses their sentiment using a Support Vector Machine combined with SMOTE and hyperparameter tuning. A total of 2,673 reviews were collected through web scraping and filtered into 2,113 final reviews, consisting of 1,275 negative and 838 positive entries. Labels were assigned automatically from the star rating, and the text passed through six preprocessing stages, including negation handling. The model was evaluated across five cumulative scenarios within a Stratified 10-Fold Cross Validation framework, where TF-IDF and SMOTE were recomputed inside every fold on the training data only to avoid data leakage. The best scenario combined preprocessing, SMOTE, N-gram (1,2), and tuning, achieving an Accuracy of 0.8316 and an F1-Score of 0.8270 with the RBF kernel, C of 10, and Gamma of 0.01. It recognises negative reviews more accurately, with a Recall of 0.9263, than positive ones at 0.6874. Complaints centre on login failures, problems after updates, and assignment file submission.

Keyword : daftarkan hingga 6 kata kunci di sini (Keyword must be typed in Italic and consist of 3-6 words or phrases)

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

1. PENDAHULUAN

Perguruan tinggi di Indonesia kini bertumpu pada perangkat lunak untuk hampir seluruh urusan akademik, dan pandemi COVID-19 mempercepat pergeseran itu. Salah satu learning management system yang paling banyak dipakai adalah EdLink buatan PT Sentra Vidya Utama atau SEVIMA, yang diperkenalkan pada tahun 2020 (SEVIMA, 2020). Sampai tahun 2023 aplikasi ini telah dipakai lebih dari 700 institusi dengan pengguna menembus tiga juta orang (Erna, 2023), dan telah diunduh lebih dari satu juta kali di Google Play Store (SEVIMA, 2026).

Angka sebesar itu membawa konsekuensi. Setiap hari pengguna menuliskan pengalamannya di kolom ulasan, kadang berupa pujian singkat, lebih sering berupa keluhan teknis. Rininda et al. (2023) mencatat keterlambatan notifikasi sebagai persoalan yang paling kerap dilaporkan, dengan proporsi keluhan mencapai 75 persen. Ulasan semacam itu sebenarnya umpan balik yang murah dan jujur, hanya saja jumlahnya menumpuk dan tidak terstruktur, sehingga membacanya satu per satu tidak realistis bagi pengembang.

Di titik inilah analisis sentimen menjadi relevan, yaitu cabang pemrosesan bahasa alami yang memetakan opini teks ke kutub positif atau negatif secara otomatis. Support Vector Machine banyak dipilih untuk tugas ini karena bekerja baik pada ruang fitur teks berdimensi tinggi (Cortes et al., 1995; Suthaharan, 2016), sebagaimana ditunjukkan Fakhri Irfani et al. (2020) pada Ruangguru dan Panjaitan dan Manurung (2022) pada Google Classroom. Namun data ulasan jarang ideal. Teksnya informal, penuh kata tidak baku dan kata negasi, sekaligus sebaran kelasnya timpang, sehingga model cenderung memihak kelas mayoritas. Chawla et al. (2002) menawarkan SMOTE untuk meredam ketimpangan itu, sedangkan optimasi parameter lazim ditempuh untuk menggali potensi model lebih jauh (Iriananda et al., 2024).

Penelitian terdahulu pada objek yang sama masih menyisakan celah. Rininda et al. (2023) menerapkan SVM pada ulasan EdLink dengan akurasi 82 persen, tetapi hanya memakai 86 data dan tanpa optimasi parameter. Mola et al. (2025) mencapai akurasi 90 persen dengan Random Forest, namun memakai skema tiga kelas pada 1.117 ulasan dan tanpa N-gram. Keduanya belum menelaah secara sistematis seberapa besar andil tiap komponen metodologi terhadap performa akhir.

Penelitian ini menutup celah tersebut dengan merancang lima skenario kumulatif, dari teks mentah sampai konfigurasi lengkap yang memuat pra-pemrosesan, SMOTE, N-gram, dan hyperparameter tuning. Rancangan bertingkat itu membuat kontribusi tiap komponen dapat diatribusikan secara langsung, bukan sekadar dilaporkan sebagai satu angka akhir. Seluruh pengujian dijalankan dalam kerangka Stratified 10-Fold Cross Validation dengan pembobotan TF-IDF dan SMOTE dihitung ulang di dalam setiap lipatan agar data uji tetap bersih dari kebocoran informasi.

2. LANDASAN TEORI

Analisis Sentimen

Analisis sentimen adalah cabang pemrosesan bahasa alami yang menentukan polaritas opini pada sebuah teks. Pada ulasan aplikasi, keluarannya biasanya disederhanakan menjadi dua kelas, yaitu puas dan tidak puas, karena batas antara ulasan netral dan keluhan ringan cenderung kabur sehingga pemisahan tiga kelas rawan menghasilkan label yang tidak konsisten.

Pembobotan TF-IDF

Teks tidak dapat langsung dicerna algoritma pembelajaran mesin. Pembobotan Term Frequency-Inverse Document Frequency mengubahnya

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

menjadi vektor numerik dengan memberi bobot tinggi pada kata yang sering muncul pada satu dokumen tetapi jarang pada dokumen lain. Haddi et al. (2013) menegaskan pembobotan semacam ini menghasilkan representasi yang jauh lebih diskriminatif dibanding sekadar menghitung frekuensi. Perhitungannya dirumuskan pada Persamaan (1) sampai (3).

$$tf(t,d) = f(t,d) / |d| \quad (1)$$
$$idf(t) = \log(N / df(t)) + 1 \quad (2)$$
$$w(t,d) = tf(t,d) \times idf(t) \quad (3)$$

Notasi f(t,d) adalah frekuensi term t pada dokumen d, |d| total term dokumen tersebut, N jumlah seluruh dokumen, dan df(t) banyaknya dokumen yang memuat term t.

SMOTE

SMOTE menyeimbangkan kelas bukan dengan menggandakan data lama, melainkan menciptakan sampel baru di sepanjang garis penghubung antara sebuah sampel minoritas dan tetangga terdekatnya (Chawla et al., 2002), sebagaimana Persamaan (4).

$$xbaru = xi + \delta\,(xj - xi),\ 0 \le \delta \le 1 \quad (4)$$

Variabel xi adalah sampel minoritas, xj salah satu tetangga terdekatnya, dan δ bilangan acak antara 0 dan 1. Karena sampel sintetis dibangkitkan dari data latih, penerapannya wajib dibatasi di dalam lipatan pelatihan agar akurasinya tidak menjadi terlalu optimistis.

Support Vector Machine

SVM mencari bidang pemisah optimal di antara dua kelas dengan memaksimalkan margin, yaitu jarak bidang tersebut terhadap titik data terluar yang disebut support vector (Suthaharan, 2016). Bidang pemisah dinyatakan pada Persamaan (5) dan fungsi keputusannya pada Persamaan (6).

$$w \cdot x + b = 0 \quad (5)$$

$$f(x) = \text{sign}(\Sigma\, \alpha i\, yi\, K(xi, x) + b) \quad (6)$$

Besar penalti pelanggaran margin dikendalikan parameter C. Untuk sebaran yang tidak terpisah secara linear, digunakan fungsi kernel yang memetakan data ke ruang berdimensi lebih tinggi. Empat kernel yang diuji adalah Linear, RBF, Polynomial, dan Sigmoid.

Metrik Evaluasi

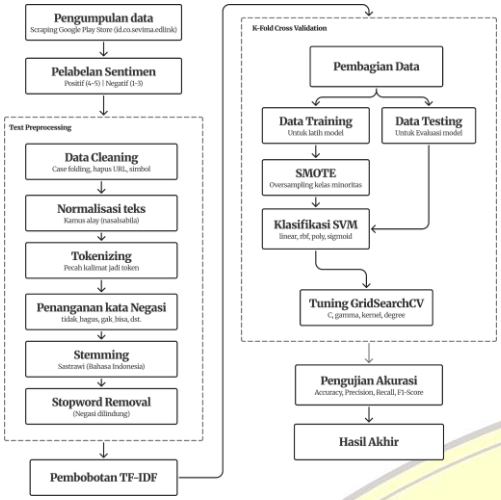
Evaluasi bertumpu pada confusion matrix yang merekam True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Powers (2011) menjelaskan empat metrik turunannya pada Persamaan (7) sampai (10). Pada data timpang, akurasi saja menyesatkan, sehingga F1-Score dipakai sebagai indikator utama karena menyeimbangkan Precision (P) dan Recall (R).

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (7)$$
$$Precision = TP / (TP + FP) \quad (8)$$
$$Recall = TP / (TP + FN) \quad (9)$$
$$F1\text{-}Score = 2 \times (P \times R) / (P + R) \quad (10)$$

3. METODOLOGI

Penelitian ini bersifat kuantitatif eksperimental dan dijalankan melalui tujuh tahapan berurutan, dari pengumpulan data sampai penarikan kesimpulan, sebagaimana ditampilkan pada Gambar 1.

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054



Gambar 1. Alur penelitian

Pengumpulan dan Pelabelan Data

Google Play Store tidak menyediakan API publik resmi untuk ekstraksi ulasan, sehingga data ditarik melalui web scraping terprogram pada Python dengan pustaka google-play-scraper. Penarikan dijalankan iteratif memakai continuation token dan berhenti otomatis ketika server tidak lagi mengembalikan ulasan baru. Target scraping adalah paket id.co.sevima.edlink dengan bahasa dan negara Indonesia, diurutkan dari yang terbaru. Dari sebelas atribut bawaan, hanya empat yang dipertahankan, yaitu userName, score, content, dan at.

Pelabelan ditetapkan otomatis dengan pendekatan rule-based berbasis rating bintang. Ulasan berating 4 dan 5 dikategorikan positif, sedangkan rating 1, 2, dan 3 dikategorikan negatif. Rating 3 sengaja dimasukkan ke kutub negatif karena nilai tengah pada penilaian aplikasi digital umumnya menandakan pengalaman yang belum memenuhi harapan. Skema ini sejalan dengan Maulana et al. (2023).

Pra-pemrosesan Teks

Enam langkah dijalankan berurutan. Pertama, data cleaning yang mencakup case folding, penghapusan URL, mention, tagar, angka, tanda baca, emoji, serta perapian spasi. Kedua, normalisasi kata tidak baku memakai Colloquial Indonesian Lexicon dari Salsabila et al. (2018), yang versi repositori publiknya memuat 4.331 pasangan kata. Ketiga, tokenisasi dengan pustaka NLTK. Keempat, penanganan kata negasi. Tujuh belas kata negasi dikenali, dan setiap kali salah satunya terdeteksi, kata tersebut digabungkan dengan token sesudahnya memakai garis bawah, misalnya tidak_bagus dan enggak_bisa. Langkah ini sengaja ditaruh sebelum stemming supaya token gabungan tidak terurai. Kelima, stemming dengan PySastrawi yang berbasis kaidah morfologi Bahasa Indonesia (Adriani et al., 2007), di mana token bernegasi hanya di-stem pada bagian setelah garis bawah. Keenam, stopword removal memakai gabungan daftar Kaggle dan PySastrawi ditambah beberapa partikel, total 756 kata setelah deduplikasi. Token bernegasi dilindungi dari penghapusan.

Ekstraksi Fitur dan Skenario

Teks bersih diubah menjadi vektor numerik memakai TfidfVectorizer dari scikit-learn dengan batas maksimum 5.000 fitur. Dua konfigurasi utama dibangun, yaitu unigram (1,1) untuk Skenario 1 sampai 3 dan gabungan unigram-bigram (1,2) untuk Skenario 4 dan 5. Lima skenario disusun kumulatif, artinya setiap skenario menambahkan tepat satu komponen baru, sehingga selisih performa dapat langsung diatribusikan pada komponen yang baru ditambahkan. Rinciannya pada Tabel 1.

Tabel 1. Lima skenario eksperimen SVM

Skenario	Konfigurasi
S1	Teks mentah, TF-IDF unigram, SVM default
S2	S1 ditambah pra-pemrosesan lengkap
S3	S2 ditambah SMOTE
S4	S3 ditambah N-gram (1,2)
S5	S4 ditambah hyperparameter tuning

P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Validasi Solang dan Optimasi Parameter

Pengujian memakai Stratified 10-Fold Cross Validation. Kohavi (1995) menunjukkan validasi silang menghasilkan estimasi yang lebih stabil dibanding pembagian data tunggal, sedangkan Prusty et al. (2022) menekankan varian stratified penting saat kelas timpang agar proporsi kelas tiap lipatan terjaga. Pembobotan TF-IDF dihitung ulang di dalam setiap lipatan hanya dari data latih, lalu data uji sekadar ditransformasi. SMOTE pun hanya menyentuh data latih. Pencarian parameter pada Skenario 5 memakai GridSearchCV yang dibungkus pipeline berisi TF-IDF, SMOTE, dan SVM, sejalan dengan strategi tuning kernel Gaussian oleh Chen et al. (2017). Ruang parameter mencakup C bernilai 0,1 sampai 100, Gamma scale, auto, 0,01, dan 0,001, empat jenis kernel, serta degree 2 sampai 4, menghasilkan 192 kombinasi dan 1.920 kali fitting. Metrik pencarian adalah f1_weighted. Seluruh eksperimen dijalankan di Google Colaboratory memakai Python 3.

4. HASIL DAN PEMBAHASAN

Proses scraping menghimpun 2.673 ulasan mentah selama empat belas batch. Pembersihan bertingkat membuang 132 ulasan yang lebih baru dari 31 Desember 2025 dan 376 ulasan duplikat, menyisakan 2.165 ulasan bersih. Distribusi ratingnya terpolarisasi di dua kutub, yaitu rating 1 bintang mendominasi dengan 796 ulasan atau 36,8 persen dan rating 5 bintang sebanyak 679 ulasan atau 31,4 persen, sementara rating menengah relatif sedikit. Kedua kutub ekstrem itu menguasai 68,2 persen dataset, dan kondisi ini justru mendukung penyederhanaan menjadi skema dua kelas. Setelah aturan pelabelan diterapkan, diperoleh 1.288 ulasan negatif atau 59,5 persen dan 877 ulasan positif atau 40,5 persen, dengan rasio ketidakseimbangan 1,47.

Hasil Pra-pemrosesan

Jejak transformasi satu contoh ulasan pada Tabel 2 memverifikasi bahwa setiap langkah bekerja sesuai fungsinya.

Tabel 2. Dokumentasi tahapan pra-pemrosesan pada satu contoh ulasan

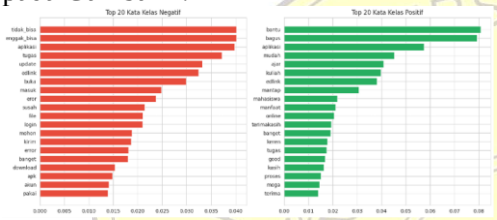
Tahap	Hasil
Teks asli	Aplikasi ini TIDAK bagus bgt, gak bisa login!! loadingnya lamaaaaa
Cleaning	aplikasi ini tidak bagus bgt gak bisa login loadingnya lamaa
Normalisasi	aplikasi ini tidak bagus banget enggak bisa login loadingnya lamaa
Tokenizing	['aplikasi', 'ini', 'tidak', 'bagus', 'banget', 'enggak', 'bisa', 'login', 'loadingnya', 'lamaa']
Negasi	['aplikasi', 'ini', 'tidak_bagus', 'banget', 'enggak_bisa', 'login', 'loadingnya', 'lamaa']
Stopword removal	['aplikasi', 'tidak_bagus', 'banget', 'enggak_bisa', 'login', 'loadingnya', 'lamaa']

Berdasarkan Tabel 2, kata tidak baku bgt berubah menjadi banget dan gak menjadi enggak. Pasangan kata tidak bagus serta enggak bisa berhasil disatukan menjadi token tunggal, lalu bertahan utuh melewati stemming dan stopword removal. Inilah bukti bahwa konteks negatif tidak hilang di tengah jalan. Setelah keenam langkah diterapkan pada seluruh dataset, sebanyak 52 ulasan berubah menjadi teks kosong dan dikeluarkan, sehingga jumlah data menyusut dari 2.165 menjadi 2.113 ulasan, terdiri atas 1.275 ulasan negatif dan 838 ulasan positif. Komposisi inilah yang menjadi dataset definitif untuk pemodelan.

https://doi.org/10.37817/ikraith-informatika.v10i2
P-ISSN : 2580-4316
E-ISSN : 2654-8054

Hasil Pembobotan TF-IDF

Matriks unigram menghasilkan 2.735 fitur, artinya seluruh kosakata unik masih tertampung di bawah batas 5.000 tanpa pemotongan. Konfigurasi bigram dan trigram langsung menyentuh plafon 5.000 fitur karena pasangan dan tripel kata memperlebar ruang fitur secara drastis. Konfigurasi bigram akhirnya dipilih sebagai masukan Skenario 4 dan 5 dengan pertimbangan keseimbangan antara kekayaan fitur dan kerapatan matriks. Dua puluh term dengan rata-rata skor TF-IDF tertinggi pada tiap kelas divisualisasikan pada Gambar 2.



Gambar2.Dua puluh term teratas berdasarkan rata-rata skor TF-IDF pada tiap kelas

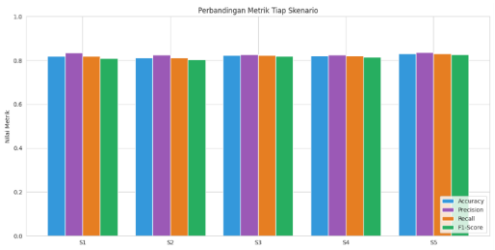
Gambar 2 memperlihatkan kontras leksikal yang tegas. Pada kelas positif, term bantu memuncaki daftar dengan skor 0,0808, disusul bagus sebesar 0,0793. Kemunculan kata mudah, ajar, kuliah, mantap, serta manfaat menunjukkan pengguna yang puas menyoroti kemudahan aplikasi dalam menopang perkuliahan daring. Pada kelas negatif, dua term gabungan negasi justru bertengger di puncak, yaitu tidak_bisa dan enggak_bisa yang sama-sama bernilai 0,0402. Temuan ini menjadi bukti empiris bahwa mekanisme penanganan negasi bekerja. Seandainya kata negasi dibiarkan berdiri sendiri, kata bisa akan tercatat sebagai term netral dan sinyal ketidakmampuan yang justru paling penting akan menguap. Kata tugas, update, buka, eror, login, susah, file, dan kirim melengkapi daftar, menandakan akar keluhan bersumber dari kendala teknis saat masuk akun, membuka aplikasi setelah pembaruan, serta mengirim berkas tugas.

Perbandingan Lima Skenario

Seluruh skenario dievaluasi memakai rata-rata sepuluh lipatan. Pada lipatan pertama, misalnya, komposisi 1.147 ulasan negatif berbanding 754 ulasan positif diseimbangkan SMOTE menjadi 1.147 berbanding 1.147, sehingga total data latih naik dari 1.901 menjadi 2.294 sampel, sementara data uji tidak disentuh. Rekapitulasi performa disajikan pada Tabel 3 dan Gambar 3.

Tabel 3. Rekapitulasi performa lima skenario SVM

Model skenario	Accuracy	Precision	Recall	F1-Score
S1: Teks mentah + SVM	0,8192	0,8346	0,8192	0,8103
S2: Pra-pemrosesan + SVM	0,8121	0,8243	0,8121	0,8035
S3: Pra-pemrosesan + SMOTE + SVM	0,8235	0,8277	0,8235	0,8188
S4: Pra-pemrosesan + SMOTE + N-gram + SVM	0,8207	0,8255	0,8207	0,8154
S5: Pra-pemrosesan + SMOTE + N-gram + Tuning	0,8316	0,8359	0,8316	0,8270



P-ISSN : 2580-4316
E-ISSN : 2654-8054 https://doi.org/10.37817/ikraith-informatika.v10i2

Gambar3.Perbandingan empat metrik evaluasi pada lima skenario SVM

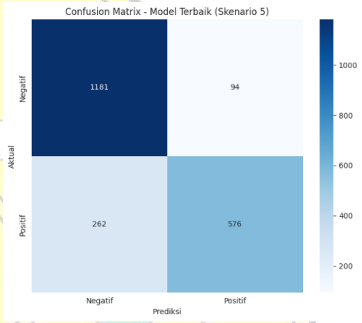
Pola pergerakan antar skenario ternyata tidak selalu menanjak, dan justru di situ letak temuan yang menarik. Skenario 2 yang menambahkan pra-pemrosesan lengkap malah turun tipis dibanding Skenario 1, dari F1-Score 0,8103 menjadi 0,8035. Stemming dan stopword removal rupanya ikut memangkas sinyal kecil yang sebelumnya masih dimanfaatkan model pada teks mentah. Fenomena semacam ini lumrah pada ulasan pendek, dan menegaskan pra-pemrosesan saja belum tentu menaikkan performa bila tidak ditopang teknik lain.

Lonjakan terbesar hadir pada Skenario 3. Penambahan SMOTE mengerek F1-Score dari 0,8035 menjadi 0,8188, naik 0,0153. Selisih itu yang terbesar antar skenario dan mengonfirmasi bahwa penyeimbangan kelas adalah komponen paling berpengaruh pada penelitian ini. Sebaliknya, Skenario 4 yang mengganti unigram menjadi n-gram (1,2) malah turun tipis ke 0,8154. Ribuan fitur pasangan kata yang ditambahkan sebagian besar jarang muncul, sehingga bobot fitur penting justru terencerkan tanpa imbalan informasi kontekstual yang berarti pada ulasan yang umumnya cuma beberapa kata.

Skenario 5 mengangkatnya kembali ke titik tertinggi, dengan Accuracy 0,8316 dan F1-Score 0,8270. Kenaikan F1-Score sebesar 0,0116 dibanding Skenario 4 membuktikan tuning berhasil menemukan kombinasi parameter yang lebih tegas dari setelan bawaan, sekaligus menebus kerugian akibat penambahan n-gram. Pelajarannya cukup jelas, menumpuk teknik tanpa penyetelan parameter tidak menjamin apa pun.

Parameter Terbaik dan Evaluasi Model

GridSearchCV menemukan kombinasi terbaik berupa kernel RBF, C sebesar 10, dan Gamma 0,01, dengan F1-Score validasi silang 0,8176. Kernel RBF menguasai seluruh sepuluh posisi teratas hasil pencarian, sesuai karakteristik matriks TF-IDF yang berdimensi tinggi. Yang patut digarisbawahi, kombinasi ini sama sekali bukan setelan bawaan scikit-learn yang memakai C sebesar 1 dan Gamma scale. Setelan bawaan itu hanya menempati peringkat ketujuh dengan F1-Score 0,8074, sehingga tuning terbukti memberi nilai tambah nyata. Confusion matrix gabungan sepuluh lipatan untuk model terbaik disajikan pada Gambar 4 dan rincian metrik per kelasnya pada Tabel 4.



Gambar3.Confusion matrix model terbaik pada Skenario 5

Tabel 4. Hasil evaluasi per kelas model terbaik

Kelas	Precision	Recall	F1-Score	Jumlah data
Negatif	0,8184	0,9263	0,8690	1.275
Positif	0,8597	0,6874	0,7639	838
Akurasi	-	-	0,8315	2.113
Weighted avg	0,8348	0,8315	0,8273	2.113

Berdasarkan Gambar 4, dari 1.275 ulasan negatif sebanyak 1.181 diprediksi benar dan 94 keliru dianggap positif. Dari 838 ulasan positif, hanya 576 yang tertangkap benar, sementara 262 sisanya justru terlempar ke kelas negatif. Ketimpangan itu tergambar jelas pada Recall, yaitu 0,9263 untuk kelas negatif berbanding 0,6874 untuk kelas positif.

Model jelas lebih tajam mengenali keluhan daripada pujian.

Menariknya, Precision kelas positif justru lebih tinggi, yakni 0,8597. Artinya, ketika model berani memberi label positif, tebakannya biasanya tepat. Model bersikap konservatif dan baru melabeli positif bila sinyalnya benar-benar kuat. Kondisi ini masuk akal mengingat kelas negatif tetap mayoritas meski SMOTE diterapkan pada data latih, sebab data uji sengaja dibiarkan dalam komposisi aslinya. Sebagian ulasan positif juga ditulis sangat pendek dan miskin kata bermuatan sentimen, sehingga sulit dibedakan dari ulasan netral yang terlanjur dilabeli negatif akibat rating 3.

Dibandingkan penelitian terdahulu pada objek yang sama, akurasi 83,15 persen di sini sedikit melampaui capaian Rininda et al. (2023) yang sebesar 82 persen, padahal korpus yang dipakai jauh lebih besar, yaitu 2.113 ulasan berbanding 86 ulasan. Mola et al. (2025) memang melaporkan angka lebih tinggi, yakni 90 persen, namun perbandingan langsung tidak sepenuhnya adil karena jumlah kelas, ukuran dataset, dan skema validasinya berbeda. Nilai pada penelitian ini diperoleh dari sepuluh lipatan tanpa kebocoran data, sehingga estimasinya cenderung konservatif dan lebih dapat dipercaya. Kontribusinya yang lebih substantif justru terletak pada terurainya andil tiap komponen, sesuatu yang belum dilakukan kedua penelitian tersebut.

5. KESIMPULAN

Analisis sentimen ulasan aplikasi EdLink berhasil diterapkan memakai Support Vector Machine yang dipadukan dengan SMOTE dan hyperparameter tuning. Dari 2.673 ulasan mentah diperoleh 2.113 ulasan final, terdiri atas 1.275 ulasan negatif dan 838 ulasan positif, dengan rasio ketidakseimbangan 1,47. Lima skenario kumulatif memperlihatkan tiap komponen memberi pengaruh berbeda. SMOTE menyumbang kenaikan terbesar dengan tambahan F1-Score 0,0153, penambahan n-gram tanpa penyetelan justru menurunkan performa secara tipis, lalu hyperparameter tuning mengembalikannya ke titik tertinggi dengan tambahan 0,0116.

Model terbaik diperoleh pada Skenario 5 dengan Accuracy 0,8316 dan F1-Score 0,8270, memakai kernel RBF, C sebesar 10, dan Gamma 0,01. Evaluasi confusion matrix menunjukkan model lebih peka menangkap ulasan negatif dengan Recall 0,9263 dibanding ulasan positif yang Recall-nya 0,6874, dengan akurasi keseluruhan 83,15 persen. Secara substansi, keluhan pengguna terkonsentrasi pada kegagalan login, kendala setelah pembaruan aplikasi, serta pengiriman berkas tugas, dan ketiganya layak menjadi prioritas perbaikan bagi pengembang.

6. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Djuanda Bogor, atas dukungan akademik selama penelitian berlangsung, serta kepada kedua dosen pembimbing atas arahan dan koreksinya.

DAFTAR PUSTAKA

Adriani, M., Asian, J., Nazief, B., Tahaghoghi, S. M. M., & Williams, H. E. (2007). Stemming Indonesian. ACM Transactions on Asian Language Information Processing, 6(4), 1–33. https://doi.org/10.1145/1316457.1316459

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16,

321–357. https://doi.org/10.1613/jair.953

Chen, G., Florero-Salinas, W., & Li, D. (2017). Simple, fast and accurate hyper-parameter tuning in Gaussian-kernel SVM. 2017 International Joint Conference on Neural Networks (IJCNN), 348–355. https://doi.org/10.1109/IJCNN.2017.7965875

Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Erna. (2023). Apa itu SEVIMA? Profil dan sejarah edtech yang sudah 2 dekade merevolusi pendidikan tinggi Indonesia. SEVIMA. https://sevima.com/apa-itu-sevima-profil-dan-sejarah-edtech-yang-sudah-2-dekade-merevolusi-pendidikan-tinggi-indonesia/

Fakhri Irfani, F., Triyanto, M., Hartanto, A. D., & Kusnawi. (2020). Analisis sentimen review aplikasi Ruangguru menggunakan algoritma Support Vector Machine. Jurnal Teknologi Informasi dan Komunikasi.

Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. Procedia Computer Science, 17, 26–32. https://doi.org/10.1016/j.procs.2013.05.005

Iriananda, S. W., Budiawan, R. W., Rahman, A. Y., & Istiadi, I. (2024). Optimasi klasifikasi sentimen komentar pengguna game bergerak menggunakan SVM, grid search dan kombinasi N-gram. Jurnal Teknologi Informasi dan Ilmu Komputer, 11(4), 743–752. https://doi.org/10.25126/jtiik.1148244

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the 14th International Joint Conference on Artificial Intelligence, 2, 1137–1143.

Maulana, R., Voutama, A., & Ridwan, T. (2023). Analisis sentimen ulasan aplikasi MyPertamina pada Google Play Store menggunakan algoritma NBC. Jurnal Teknologi Terpadu, 9(1), 42–48. https://doi.org/10.54914/jtt.v9i1.609

Mola, S. A. S., Polly, D. P. N., & Rumlaklak, N. D. (2025). Sentiment analysis on user reviews of the Edlink application using the Random Forest Classifier method. Jurnal Sisfotek Global, 15(1), 20–27. https://doi.org/10.38101/sisfotek.v15i1.15788

Panjaitan, N., & Manurung, N. (2022). Sentiment analysis of Google Classroom application using Support Vector Machine. 2022 6th International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM), 117–120. https://doi.org/10.1109/ELTICOM57747.2022.10038262

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63.

Pratama, M. R., Wardani, R. R. K., Hapsari, S. S., & Dewi, M. A. (2023). User experience analysis on the Edlink mobile application using usability testing method. 2023 8th International Conference on Business and Industrial Research (ICBIR), 942–947.

https://doi.org/10.1109/ICBIR57571.2023.10147656

Prusty, S., Patnaik, S., & Dash, S. K. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. Frontiers in Nanotechnology, 4, 972421. https://doi.org/10.3389/fnano.2022.972421

Rininda, G., Hartami Santi, I., & Kirom, S. (2023). Penerapan SVM dalam analisis sentimen pada EdLink menggunakan pengujian confusion matrix. JATI (Jurnal Mahasiswa Teknik Informatika), 7(5), 3335–3342. https://doi.org/10.36040/jati.v7i5.7420

Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. 2018 International Conference on Asian Language Processing (IALP), 226–229. https://doi.org/10.1109/IALP.2018.8629151

SEVIMA. (2020). SEVIMA EdLink, aplikasi LMS terbaik karya anak bangsa. PT Sentra Vidya Utama. https://sevima.com/sevima-edlink-aplikasi-lms-terbaik-karya-anak-bangsa/

SEVIMA. (2026). SEVIMA EdLink [Aplikasi di Google Play]. Google Play Store. https://play.google.com/store/apps/details?id=id.co.sevima.edlink

Sulistiana, & Aziz Muslim, M. (2020). Support Vector Machine (SVM) optimization using grid search and unigram to improve e-commerce review accuracy. Journal of Soft Computing Exploration, 1(1). https://doi.org/10.52465/joscex.v1i1.3

Suthaharan, S. (2016). Support Vector Machine. In Machine learning models and algorithms for big data classification (Vol. 36, pp. 207–235). Springer. https://doi.org/10.1007/978-1-4899-7641-3_9