

Analisis Sentimen Ulasan IKD Berbasis Lexicon dan SMOTE pada Naive Bayes, SVM, dan Random Forest

¹Adisha Dhia Bimantara, ²Bambang Adiwino

^{1,2}Sistem Informasi, ISB Atma Luhur, Pangkalpinang, Indonesia

*Korespondensi: 2222500084@mahasiswa.atmaluhur.ac.id

Submit : 19 April 2026 | Diterima : 18 Mei 2026 | Terbit : 22 Mei 2026

ABSTRACT

The rapid adoption of mobile-based public services in Indonesia has driven increased utilization of the Identitas Kependudukan Digital (IKD) application for population administration. As the user base expands, review data from the Google Play Store has grown substantially, reflecting varied user experiences encompassing service satisfaction, usability perceptions, and technical obstacles. Processing this volume of unstructured text manually is impractical, necessitating an automated computational approach to opinion mining. This study investigates opinion polarity in IKD user reviews through a lexicon-based labeling strategy and evaluates the classification capacity of Naive Bayes, Support Vector Machine (SVM), and Random Forest algorithms both prior to and following the application of Synthetic Minority Oversampling Technique (SMOTE). A total of 4,896 reviews were gathered from Google Play Store, with 3,708 categorized as negative and 1,188 as positive sentiments. The pipeline encompassed text preprocessing, automatic lexicon-driven labeling, TF-IDF vectorization, an 80:20 train-test partition, and multi-metric model evaluation. Experimental findings confirm that SMOTE effectively addresses class imbalance by bolstering the representation of the minority class. Among all configurations tested, SVM combined with SMOTE yielded superior results, recording 98.98% accuracy, 98.29% precision, 97.46% recall, and 97.87% F1-score. These outcomes demonstrate that integrating lexicon-based annotation, TF-IDF feature weighting, and SMOTE resampling constitutes an effective pipeline for sentiment classification in digital government applications, offering a scalable mechanism for evidence-based public service quality evaluation.

Keywords: IKD; lexicon; machine learning; sentiment analysis; SMOTE

ABSTRAK

Penerapan teknologi digital pada sektor administrasi publik mendorong meluasnya pemakaian aplikasi Identitas Kependudukan Digital (IKD) sebagai kanal pelayanan kependudukan berbasis perangkat seluler. Pertumbuhan basis pengguna aplikasi ini turut menghasilkan akumulasi ulasan pada Google Play Store yang memuat ragam perspektif pengguna mencakup kepuasan layanan, kemudahan antarmuka, serta hambatan teknis yang ditemui. Besarnya volume data tekstual ini membuat proses kajian manual menjadi tidak efektif sehingga dibutuhkan metode berbasis komputasi untuk mengekstraksi opini pengguna secara otomatis. Penelitian ini bertujuan mengklasifikasikan polaritas sentimen ulasan pengguna aplikasi IKD dengan memanfaatkan kamus sentimen berbasis leksikon sebagai mekanisme pelabelan otomatis, sekaligus membandingkan kinerja algoritma Naive Bayes, Support Vector Machine (SVM), dan Random Forest sebelum dan setelah penerapan Synthetic Minority Oversampling Technique (SMOTE). Data yang digunakan mencakup 4.896 ulasan yang dikumpulkan dari Google Play Store, terdiri atas 3.708 ulasan bersentimen negatif dan 1.188 ulasan bersentimen positif. Alur kerja penelitian meliputi praproses teks, pelabelan otomatis berbasis leksikon, pembobotan fitur TF-IDF, pembagian data dengan rasio 80:20, serta evaluasi model menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil eksperimen memperlihatkan bahwa SMOTE mampu memperbaiki ketimpangan distribusi kelas secara signifikan. Model SVM yang dikombinasikan dengan SMOTE menghasilkan kinerja terbaik dengan akurasi 98,98%, presisi 98,29%, recall 97,46%, dan F1-score 97,87%. Temuan ini mengindikasikan bahwa integrasi pelabelan berbasis leksikon, pembobotan TF-IDF, dan penyeimbangan SMOTE merupakan pipeline analisis sentimen yang efektif untuk mendukung pemantauan kualitas layanan publik digital berbasis opini pengguna.

Kata Kunci: analisis sentimen; IKD; lexicon; machine learning; SMOTE

PENDAHULUAN

Penetrasi teknologi digital pada ranah pemerintahan mendorong institusi publik di Indonesia untuk menyediakan layanan administrasi kependudukan berbasis aplikasi seluler guna memperluas keterjangkauan dan mempercepat prosedur pelayanan kepada warga negara (Tanggaraeni & Sitokdana, 2022). Salah satu manifestasi nyata dari agenda transformasi ini adalah aplikasi Identitas Kependudukan Digital (IKD), yang memberikan kemudahan akses dokumen kependudukan secara elektronik tanpa keharusan pengguna hadir secara fisik di kantor dinas (Saputra, 2023). Kehadiran platform layanan publik digital semacam ini tidak hanya menyederhanakan prosedur administratif, melainkan turut membuka kanal komunikasi dua arah antara pengguna dan penyelenggara layanan melalui mekanisme ulasan pada marketplace digital seperti Google Play Store (Cahyani & Prasetyaningrum, 2026). Kumpulan ulasan yang terhimpun pada platform tersebut memiliki nilai strategis yang tinggi, sebab di dalamnya terkandung representasi nyata terkait pengalaman pengguna, tingkat penerimaan fitur, kepuasan layanan, serta hambatan teknis yang ditemui dalam aktivitas penggunaan sehari-hari (Idris & Mussalimun, 2024).

Data ulasan pada Google Play Store hadir dalam format tidak terstruktur, tumbuh secara dinamis, dan terakumulasi dalam jumlah yang besar seiring bertambahnya basis pengguna aplikasi. Karakteristik tersebut menjadikan proses kajian manual terhadap opini pengguna menjadi tidak praktis, terutama ketika pemerintah membutuhkan umpan balik secara cepat dan berkelanjutan untuk keperluan evaluasi layanan (Xu et al., 2023). Dalam kerangka tata kelola layanan publik berbasis digital, pemantauan opini pengguna secara otomatis berperan sebagai fondasi pengambilan keputusan berbasis data, perbaikan fitur aplikasi, serta peningkatan kualitas layanan secara iteratif (Verma, 2022). Kondisi inilah yang melatarbelakangi kebutuhan terhadap metode analisis sentimen komputasional sebagai solusi untuk mengekstraksi pola opini dari data teks secara efisien, terukur, dan dapat direplikasi (Shaik et al., 2023).

Analisis sentimen merupakan subdisiplin dalam Natural Language Processing (NLP) dan text mining yang secara khusus berfokus pada identifikasi kecenderungan emosi atau polaritas opini dari konten teks ke dalam kategori positif, negatif, maupun netral (Samanmali & Rupasingha, 2024). Metode ini telah banyak digunakan dalam berbagai studi ulasan aplikasi digital karena mampu mengonversi data tekstual tak terstruktur menjadi informasi kuantitatif yang dapat diinterpretasikan secara sistematis (Maulana et al., 2024). Salah satu pendekatan yang populer digunakan adalah metode berbasis leksikon (lexicon-based), yang bekerja dengan mencocokkan token teks terhadap kamus sentimen berisi entri kata dengan bobot polaritas positif atau negatif (Sharma & Ghose, 2021). Kelebihan pendekatan ini terletak pada kemampuannya menghasilkan label sentimen secara otomatis tanpa memerlukan data berlabel hasil anotasi manual, sehingga sangat cocok diterapkan pada dataset berukuran besar di mana keterlibatan anotator manusia menjadi kendala sumber daya (Lin & Nuha, 2023).

Sejumlah kajian empiris sebelumnya telah mengeksplorasi penggunaan berbagai algoritma supervised learning untuk klasifikasi sentimen ulasan aplikasi digital. Algoritma Naive Bayes terbukti memberikan kinerja yang memadai pada data ulasan aplikasi e-government di Google Play Store berkat keunggulannya dalam hal efisiensi komputasi dan kemampuan bekerja secara efektif di ruang fitur berdimensi tinggi (Hokijuliandy et al., 2023). Sementara itu, Support Vector Machine (SVM) dikenal unggul dalam tugas pemisahan kelas teks berkat kemampuannya menemukan hyperplane optimal yang memaksimalkan margin antar kelas dalam ruang fitur berdimensi tinggi (Purohit et al., 2024). Di sisi lain, Random Forest yang merupakan metode ensemble berbasis banyak pohon keputusan juga menunjukkan performa kompetitif berkat kemampuannya memodelkan pola nonlinier dalam data teks melalui mekanisme voting mayoritas (Sun et al., 2020).

Meski begitu, salah satu persoalan mendasar yang kerap dijumpai pada dataset ulasan aplikasi adalah ketimpangan distribusi kelas antara ulasan positif dan negatif. Ketika proporsi antar kelas tidak seimbang secara signifikan, model klasifikasi cenderung berpihak pada kelas dengan sampel lebih banyak, sehingga kemampuan mendeteksi kelas minoritas menurun tajam yang tercermin dari rendahnya nilai recall meski akurasi global tampak tinggi (Santhosh Kumar et al., 2025). Teknik Synthetic Minority Oversampling Technique (SMOTE) dirancang sebagai solusi atas kondisi tersebut, yakni dengan membangkitkan sampel sintesis pada kelas minoritas

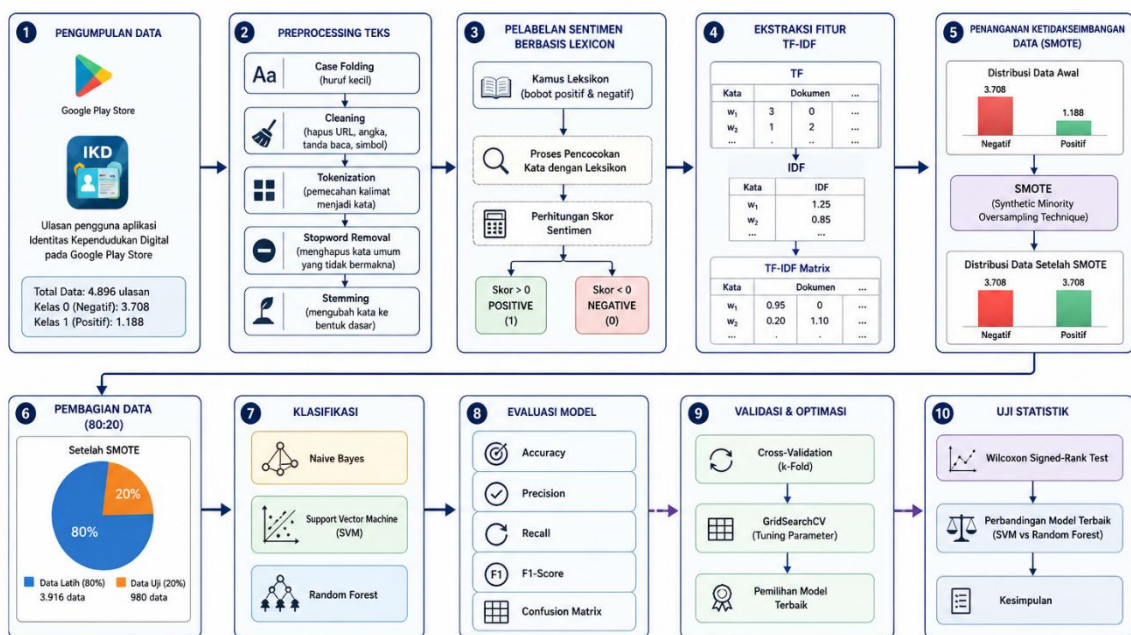
berdasarkan interpolasi linier antara titik data yang berdekatan di ruang fitur, alih-alih sekadar menduplikasi data yang sudah ada (Cahya et al., 2023). Berbagai penelitian menunjukkan bahwa integrasi SMOTE sebelum proses pelatihan model secara konsisten mampu meningkatkan kemampuan generalisasi model terhadap kelas yang kurang terwakili (Faridzi et al., 2023).

Dari eksplorasi awal terhadap data yang digunakan, ulasan pengguna aplikasi IKD yang berhasil dikumpulkan berjumlah 4.896 entri dengan komposisi 3.708 ulasan bersentimen negatif dan 1.188 bersentimen positif, sehingga rasio ketidakseimbangan kelas mencapai sekitar 3:1. Kondisi distribusi ini berpotensi menurunkan performa model jika tidak ditangani secara eksplisit sebelum tahap pelatihan (Basri et al., 2026). Selain itu, penelitian yang secara khusus mengkaji sentimen ulasan aplikasi layanan kependudukan digital di Indonesia, terutama yang mengintegrasikan pelabelan otomatis berbasis leksikon, penyeimbangan data menggunakan SMOTE, dan evaluasi komparatif multi-algoritma secara simultan, masih terbilang jarang dilakukan (Setiawan, 2024). Kesenjangan ini menegaskan perlunya studi yang lebih komprehensif, yang tidak hanya menghasilkan model dengan performa tinggi tetapi juga memberikan kontribusi metodologis bagi pengembangan sistem pemantauan kualitas layanan publik digital berbasis opini masyarakat (Kartini et al., 2023).

Berdasarkan latar belakang tersebut, penelitian ini dirancang untuk mengklasifikasikan polaritas sentimen ulasan pengguna aplikasi IKD menggunakan kamus sentimen berbasis leksikon sebagai mekanisme pelabelan otomatis, serta membandingkan kinerja tiga algoritma klasifikasi yaitu Naive Bayes, Support Vector Machine, dan Random Forest pada dua kondisi eksperimen, yakni sebelum dan setelah penerapan SMOTE. Alur penelitian mencakup tahap pra-proses teks, pelabelan sentimen berbasis leksikon, pembobotan fitur menggunakan TF-IDF, partisi data dengan skema 80:20, dan evaluasi model menggunakan metrik accuracy, precision, recall, serta F1-score. Melalui desain ini, penelitian diharapkan berkontribusi secara metodologis dalam pengembangan sistem analisis sentimen ulasan aplikasi layanan publik digital, sekaligus menghasilkan rekomendasi praktis untuk evaluasi kualitas aplikasi IKD dari sudut pandang pengalaman pengguna.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis text mining untuk mengidentifikasi polaritas opini pengguna terhadap aplikasi Identitas Kependudukan Digital pada Google Play Store. Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data, preprocessing, pelabelan sentimen, ekstraksi fitur, penanganan ketidakseimbangan kelas, pemodelan klasifikasi, hingga evaluasi performa model.



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Sumber data pada penelitian ini adalah ulasan pengguna aplikasi Identitas Kependudukan Digital (IKD) yang dikumpulkan dari platform distribusi aplikasi Google Play Store melalui teknik web scraping (Tanggraeni & Sitokdana, 2022). Keseluruhan data yang berhasil dihimpun berjumlah 4.896 ulasan yang merepresentasikan beragam pengalaman, penilaian, dan opini pengguna terhadap fungsionalitas serta kualitas layanan aplikasi.

Preprocessing Teks

Praproses teks merupakan fase kritis dalam pipeline analisis sentimen berbasis NLP yang bertujuan meningkatkan kualitas data masukan sebelum proses klasifikasi dilakukan (Semary et al., 2024). Penerapan teknik praproses yang tepat terbukti mampu mereduksi noise secara substansial serta menyeragamkan representasi kata sehingga berpengaruh positif terhadap performa model (Liang et al., 2023). Serangkaian teknik praproses yang diterapkan dalam penelitian ini adalah sebagai berikut:

1. **Case Folding**
Tahap ini, Seluruh huruf pada teks diubah menjadi huruf kecil untuk menyeragamkan representasi kata
2. **Cleaning**
Tahap ini menghapus karakter yang tidak relevan seperti URL, angka, tanda baca, simbol khusus, dan karakter non-alfabet.
3. **Tokenization**
Tahap ini, setiap teks dibagi menjadi bagian-bagian kecil, dikenal sebagai token. Token bisa berupa kata, rangkaian kata, atau kalimat
4. **Stopword Removal**
Tahap ini melakukan penghapusan kata-kata yang sering muncul dan memberikan sedikit nilai semantik
5. **Stemming**
Tahap ini mengambil kata-kata dan mengubahnya menjadi kata-kata dasar dengan cara menghapus imbuhan atau bagian belakang pada kata tersebut.

Pelabelan Sentimen Berbasis Lexicon

Pelabelan sentimen pada penelitian ini menggunakan pendekatan lexicon-based dengan memanfaatkan kamus sentimen InSet Lexicon yang merupakan kamus sentimen khusus bahasa Indonesia berisi ribuan kata positif dan negatif dengan bobot polaritas tertentu (Sharma & Ghose, 2021). Setiap token hasil praproses dicocokkan dengan entri kamus untuk memperoleh nilai skor polaritasnya, yang selanjutnya diakumulasikan menjadi skor sentimen tingkat dokumen (Lin & Nuha, 2023). Jika total skor bernilai positif, ulasan diberi label positif, sementara jika bernilai negatif, ulasan diberi label negatif. Pendekatan ini dipilih karena mampu menghasilkan label secara otomatis tanpa proses anotasi manual, sehingga lebih efisien untuk dataset dalam jumlah besar.

Ekstraksi Fitur Menggunakan TF-IDF

Usai pelabelan sentimen, representasi numerik dari data teks dihasilkan melalui metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini mengukur relevansi suatu kata dalam suatu dokumen secara relatif terhadap keseluruhan korpus, sehingga kata yang sering muncul di dokumen tertentu namun jarang di dokumen lain mendapat bobot yang lebih tinggi (Elistiana et al., 2023). Bobot TF-IDF dirumuskan sebagai $TF(t,d) \times \log(N/df(t))$, di mana $TF(t,d)$ adalah frekuensi kemunculan term t dalam dokumen d , N adalah jumlah total dokumen, dan $df(t)$ adalah jumlah dokumen yang mengandung term t (Bordoloi & Biswas, 2023). Pendekatan pembobotan ini telah terbukti efektif dalam berbagai tugas klasifikasi teks termasuk analisis sentimen karena mampu menonjolkan kata yang benar-benar diskriminatif antar kelas (Semary et al., 2024). Vektor TF-IDF yang dihasilkan selanjutnya digunakan sebagai representasi fitur masukan pada seluruh model klasifikasi.

Penanganan Ketidakseimbangan Data Menggunakan SMOTE

Ketimpangan distribusi kelas yang terdapat pada dataset penelitian ini berpotensi mendistorsi hasil pembelajaran model, di mana model cenderung mengoptimalkan prediksi untuk kelas mayoritas dengan mengorbankan kelas minoritas (Santhosh Kumar et al., 2025). Untuk menangani kondisi tersebut, penelitian ini menerapkan Synthetic Minority Oversampling

Technique (SMOTE) sebagai strategi penyeimbangan data (Cahya et al., 2023). SMOTE bekerja dengan membangkitkan sampel sintesis baru pada kelas minoritas berdasarkan kedekatan tetangga terdekat. Pendekatan ini bertujuan meningkatkan representasi kelas minoritas tanpa melakukan duplikasi data secara langsung. Pada penelitian ini, eksperimen dilakukan pada dua skenario, yaitu Tanpa SMOTE dan Dengan SMOTE. Perbandingan kedua skenario digunakan untuk mengevaluasi pengaruh SMOTE terhadap performa model.

Pembagian Data

Partisi data pada penelitian ini dilaksanakan menggunakan rasio 80:20 yang merupakan skema pembagian data yang umum digunakan dalam penelitian klasifikasi teks (Wajdi et al., 2024). Dari total 980 sampel data uji, sebanyak 744 merupakan data bersentimen negatif dan 236 merupakan data bersentimen positif. Data latih digunakan untuk membangun model klasifikasi, sementara data uji berfungsi sebagai basis evaluasi performa model pada data yang belum pernah diamati sebelumnya.

Algoritma Klasifikasi

Tiga algoritma klasifikasi dipilih dalam penelitian ini berdasarkan relevansinya dalam tugas klasifikasi teks sentimen. Pertama, Naive Bayes, yaitu algoritma berbasis probabilitas yang menerapkan Teorema Bayes dengan asumsi independensi kondisional antar fitur, dikenal efisien secara komputasional dan handal pada data teks berdimensi tinggi (Hokijulandy et al., 2023). Kedua, Support Vector Machine (SVM), yang bekerja dengan menemukan hyperplane pemisah optimal yang memaksimalkan margin antara dua kelas dalam ruang fitur berdimensi tinggi, menunjukkan kemampuan generalisasi yang kuat pada tugas klasifikasi teks (Purohit et al., 2024). Ketiga, Random Forest, yaitu metode ensemble yang membangun sejumlah pohon keputusan secara paralel menggunakan bootstrap sampling dan menggabungkan prediksi melalui mekanisme majority voting, terbukti tangguh dalam menangani fitur dengan dimensi tinggi dan kompleksitas data nonlinier (Ferdiana, 2022).

Evaluasi Model

Kinerja model klasifikasi pada penelitian ini dievaluasi menggunakan empat metrik standar yang diturunkan dari confusion matrix, yaitu accuracy, precision, recall, dan F1-score (Elistiana et al., 2023). Pemilihan keempat metrik ini dinilai komprehensif untuk mengukur performa model pada dataset tidak seimbang karena mampu menangkap dimensi evaluasi yang berbeda, mulai dari akurasi global hingga kemampuan model dalam mendeteksi kelas minoritas secara spesifik (Ashari et al., 2025).

Accuracy

Accuracy menunjukkan proporsi keseluruhan prediksi yang benar terhadap seluruh data.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision

Precision menunjukkan tingkat ketepatan model dalam memprediksi kelas positif.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall

Recall menunjukkan kemampuan model dalam menemukan seluruh data positif yang sebenarnya

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1-score

F1-score merupakan rata-rata harmonik antara precision dan recall.

(4)

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

HASIL DAN PEMBAHASAN

Distribusi Dataset

Dataset penelitian terdiri atas 4.896 ulasan pengguna aplikasi Identitas Kependudukan Digital (IKD) pada Google Play Store. Berdasarkan proses pelabelan sentimen berbasis lexicon, data terbagi menjadi 3.708 ulasan negatif dan 1.188 ulasan positif.

Tabel 1. Distribusi Dataset

Kelas	Jumlah
Negatif (0)	3.708
Positif (1)	1.188

Pembagian Data

Dataset dibagi menggunakan skema 80:20. Sebanyak 80% data digunakan sebagai data pelatihan dan 20% sebagai data pengujian. Berdasarkan pembagian tersebut, data uji berjumlah 980 data yang terdiri atas 744 data negatif dan 236 data positif

Tabel 2. Pembagian Data

Kelas	Jumlah
Negatif (0)	744
Positif (1)	236
Total	980

Dari tabel 2, Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengevaluasi performa model pada data yang belum pernah dilihat sebelumnya.

Hasil Klasifikasi Tanpa SMOTE

Pengujian awal dilakukan tanpa penerapan SMOTE. Hasil eksperimen menunjukkan bahwa algoritma Support Vector Machine (SVM) memperoleh performa terbaik dengan akurasi 98,06%, precision 98,22%, recall 93,64%, dan F1-score 95,88%. Sementara itu, Naive Bayes menghasilkan akurasi 88,06% dengan recall 51,69%, sedangkan Random Forest memperoleh akurasi 83,98% dengan recall 34,32%.

Tabel 3. Hasil Klasifikasi Tanpa SMOTE

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.8806	0.9760	0.5169	0.6759
SVM	0.9806	0.9822	0.9364	0.9588
Random Forest	0.8398	0.9759	0.3432	0.5078

Nilai recall yang rendah pada Naive Bayes dan Random Forest menunjukkan bahwa kedua model mengalami kesulitan dalam mendeteksi kelas positif sebagai kelas minoritas. Meskipun precision kedua model relatif tinggi, rendahnya recall mengindikasikan bahwa sebagian besar data positif masih terklasifikasi ke dalam kelas negatif. Hal tersebut memperlihatkan pengaruh ketidakseimbangan kelas terhadap performa model.

Hasil Klasifikasi Dengan SMOTE

Pada tahap berikutnya dilakukan eksperimen dengan menerapkan Synthetic Minority Oversampling Technique (SMOTE). Hasil menunjukkan peningkatan performa pada seluruh model. Naive Bayes memperoleh akurasi 89,90%, precision 75,09%, recall 86,86%, dan F1-score 80,55%. Random Forest memperoleh akurasi 96,94%, precision 97,25%, recall 89,83%,

dan F1-score 93,39%. Model SVM tetap menjadi model terbaik dengan akurasi 98,98%, precision 98,29%, recall 97,46%, dan F1-score 97,87%.

Tabel 4. Hasil Klasifikasi Dengan SMOTE

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.8990	0.7509	0.8686	0.8055
SVM	0.9898	0.9829	0.9746	0.9787
Random Forest	0.9694	0.9725	0.8983	0.9339

Penerapan SMOTE terbukti meningkatkan kemampuan model dalam mengenali kelas positif. Peningkatan paling signifikan terjadi pada Random Forest, di mana recall meningkat dari 34,32% menjadi 89,83%. Hal ini menunjukkan bahwa distribusi data yang lebih seimbang mampu memperbaiki kemampuan model dalam mendeteksi kelas minoritas.

Ablation Study

Analisis ablation dilakukan untuk mengevaluasi kontribusi SMOTE terhadap performa klasifikasi. Pada skenario TF-IDF saja, model menghasilkan akurasi 98,06%, recall 93,64%, dan F1-score 95,88%. Setelah penerapan TF-IDF + SMOTE, akurasi meningkat menjadi 98,98%, recall menjadi 97,46%, dan F1-score menjadi 97,87%.

Tabel 5. Ablation Study

Skenario	Accuracy	Precision	Recall	F1-Score
TF-IDF Only	0.9806	0.9822	0.9364	0.9588
TF-IDF + SMOTE	0.9898	0.9829	0.9746	0.9787

Hasil tersebut menunjukkan bahwa SMOTE memberikan kontribusi positif terhadap peningkatan performa model, khususnya pada metrik recall dan F1-score. Dengan demikian, penggunaan SMOTE efektif dalam mengurangi bias model terhadap kelas mayoritas.

Cross-Validation

Evaluasi menggunakan cross-validation menunjukkan bahwa SVM memiliki nilai rata-rata F1-score tertinggi sebesar 0,9418 dengan standar deviasi 0,0106. Naive Bayes memperoleh mean F1 sebesar 0,6954 dengan standar deviasi 0,0146, sedangkan Random Forest memperoleh mean F1 sebesar 0,4665 dengan standar deviasi 0,0332.

Tabel 6. Hasil Cross-Validation

Model	Mean F1	Std
Naive Bayes	0.6954	0.0146
SVM	0.9418	0.0106
Random Forest	0.4665	0.0332

Nilai rata-rata F1-score yang tinggi dan standar deviasi yang rendah menunjukkan bahwa SVM memiliki performa yang lebih stabil dan konsisten dibandingkan model lainnya.

Pembahasan

Temuan penelitian secara keseluruhan mengkonfirmasi bahwa ketimpangan distribusi kelas

memberikan dampak nyata terhadap performa model klasifikasi, utamanya pada kemampuan model dalam mengenali kelas minoritas. Hal ini sejalan dengan temuan (Kedas et al., 2022) yang menyatakan bahwa dataset tidak seimbang secara konsisten menghasilkan bias prediksi terhadap kelas dominan. Pada kondisi eksperimen tanpa SMOTE, Naive Bayes mencatat akurasi 88,06%, precision 97,60%, recall 51,69%, dan F1-score 67,59%. Meskipun nilai precision yang tinggi mengindikasikan ketepatan prediksi positif yang baik, rendahnya recall menegaskan bahwa model gagal menangkap sebagian besar sampel positif. Pola serupa teramati pada Random Forest dengan akurasi 83,98%, precision 97,59%, recall 34,32%, dan F1-score 50,78%, yang menunjukkan tingkat false negative yang sangat tinggi akibat dominasi kelas negatif dalam data latih. Kondisi ini konsisten dengan pola yang dilaporkan dalam studi analisis sentimen pada dataset tidak seimbang, di mana akurasi keseluruhan yang tinggi tidak selalu mencerminkan kemampuan model yang sebenarnya dalam mendeteksi kelas minoritas (Kedas et al., 2022).

Pada kondisi yang sama, Support Vector Machine (SVM) menunjukkan performa paling baik dengan akurasi 98,06%, precision 98,22%, recall 93,64%, dan F1-score 95,88%. Hasil ini menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam memisahkan dua kelas sentimen pada data teks berdimensi tinggi dibandingkan Naive Bayes maupun Random Forest.

Setelah penerapan SMOTE, seluruh model mengalami peningkatan performa yang terukur, mengkonfirmasi efektivitas teknik oversampling ini dalam mengatasi bias akibat ketidakseimbangan kelas. Naive Bayes memperlihatkan lonjakan recall dari 51,69% menjadi 86,86% dan F1-score dari 67,59% menjadi 80,55%. Peningkatan yang lebih dramatis terjadi pada Random Forest, yaitu recall melonjak dari 34,32% menjadi 89,83% dan F1-score dari 50,78% menjadi 93,39%. Peningkatan signifikan pada kedua model ini mengindikasikan bahwa keduanya sangat sensitif terhadap distribusi data latih, sehingga penyeimbangan kelas melalui SMOTE memberikan dampak yang lebih besar dibandingkan pada SVM. Temuan ini selaras dengan hasil studi (Faridzi et al., 2023) yang melaporkan bahwa oversampling secara konsisten meningkatkan kemampuan deteksi kelas minoritas pada dataset ulasan tidak seimbang.

SVM dengan SMOTE mempertahankan posisinya sebagai model terbaik dengan akurasi 98,98%, precision 98,29%, recall 97,46%, dan F1-score 97,87%. Secara komparatif terhadap skenario tanpa SMOTE, terjadi peningkatan akurasi sebesar 0,92 poin persentase, recall meningkat 3,82 poin persentase, dan F1-score meningkat 1,99 poin persentase. Capaian ini menunjukkan bahwa kombinasi representasi fitur TF-IDF dan penyeimbangan data melalui SMOTE tidak hanya memperkuat representasi kelas minoritas, tetapi juga meningkatkan kualitas batas keputusan yang dibangun oleh SVM. Hasil ini mendukung temuan (Semary et al., 2024) yang membuktikan superioritas SVM dengan TF-IDF pada tugas klasifikasi sentimen, sekaligus menguatkan argumen bahwa penanganan imbalance secara eksplisit memberikan kontribusi nyata terhadap peningkatan kinerja model klasifikasi teks (Cahyani & Prasetyaningrum, 2026).

KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan lexicon-based, TF-IDF, dan SMOTE efektif untuk analisis sentimen ulasan pengguna aplikasi Identitas Kependudukan Digital pada Google Play Store. Pada skenario tanpa SMOTE, Support Vector Machine (SVM) memberikan performa terbaik dengan akurasi 98,06% dan F1-score 95,88%. Setelah penerapan SMOTE, performa seluruh model meningkat, dengan SVM tetap menjadi model terbaik yang memperoleh akurasi 98,98%, precision 98,29%, recall 97,46%, dan F1-score 97,87%. Hasil tersebut menunjukkan bahwa penerapan SMOTE mampu meningkatkan kemampuan model dalam mengenali kelas minoritas. Dengan demikian, kombinasi lexicon-based, TF-IDF, dan SVM berbasis SMOTE terbukti efektif untuk klasifikasi sentimen ulasan aplikasi layanan publik digital.

REFERENSI

- Ashari, P., Mutiah, N., & Prawira, D. (2025). *Perbandingan Kinerja Model Deep Learning BERT Dan GPT Dalam Analisis Sentimen Komentar Video Youtube (Studi Kasus : Film Dirty Vote)*. 11(1), 66–75.
- Basri, H., Purwanti, W. N., & Alparisi, I. (2026). Evaluating SMOTE Performance For Imbalanced Multi-Label Sentiment Classification In MLSE Usability Testing Of Mobile App Reviews. *Jurnal Teknik Informatika (JUTIF)*, 7(2), 937–947. <https://doi.org/10.52436/1.Jutif.2026.7.2.5351>
- Bordoloi, M., & Biswas, S. K. (2023). Sentiment Analysis: A Survey On Design Framework,

- Applications And Future Scopes. *Artificial Intelligence Review*, 56(4), 2973–3009. <https://doi.org/10.1007/S10462-022-10144-1>
- Cahya, L. D., Luthfiarta, A., Krisna, J. I. T., Winarno, S., & Nugraha, A. (2023). Improving Multi-Label Classification Performance On Imbalanced Datasets Through SMOTE Technique And Data Augmentation Using Indobert Model. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(3), 290–298. <https://doi.org/10.25077/TEKNOSI.V9i3.2023.290-298>
- Cahyani, R. D., & Prasetyaningrum, P. T. (2026). Sentiment Analysis Of User Reviews For AI Applications: Evaluating SVM, Logistic Regression, And Random Forest. *Journal Of Information Systems And Informatics*, 8(1), 1–27. <https://doi.org/10.63158/Journalisi.V8i1.1366>
- Elistiana, K. M., Kusuma, B. A., Subarkah, P., & Rozaq, H. A. A. (2023). IMPROVEMENT OF NAIVE BAYES ALGORITHM IN SENTIMENT ANALYSIS OF SHOPEE APPLICATION REVIEWS ON GOOGLE PLAY STORE. *Jurnal Teknik Informatika (JUTIF)*, 4(6), 1431–1436. <https://doi.org/10.52436/1.Jutif.2023.4.6.1486>
- Faridzi, Z. A., Pramesti, D., & Fa'rifah, R. Y. (2023). A Comparison Of Oversampling And Undersampling Methods In Sentiment Analysis Regarding Indonesia Fuel Price Increase Using Support Vector Machine. *2023 International Conference On Advancement In Data Science, E-Learning And Information System (ICADEIS)*, 1–6. <https://doi.org/10.1109/ICADEIS58666.2023.10270851>
- Ferdiana, P. (2022). Sentiment Analysis Using Random Forest Algorithm: Online Social Media Based. *International Journal Of Recent Technology And Engineering*, 8(5), 1–8. <https://doi.org/10.35940/Ijrte.E5293.018520>
- Hokijuliandy, E., Napitupulu, H., & Firdaniza. (2023). Application Of SVM And Chi-Square Feature Selection For Sentiment Analysis Of Indonesia's National Health Insurance Mobile Application. *Mathematics*, 11(17). <https://doi.org/10.3390/Math11173765>
- Idris, M., & Mussalimun. (2024). Sentiment Analysis Of Google Play Store Reviews Using Support Vector Machines. *International Journal Of Applied Information Systems (IJ AIS)*, 12(42), 48–53. <https://doi.org/10.5120/Ijais2023451957>
- Kartini, Darmawan, A. K., Syah, R. I., & Makruf, M. (2023). Sentiment Analysis Of Social Media For Indonesian M-Health Pedulilindungi Mobile-Apps(PLMA) With Lexicon-Based And Support Vector Machine Approach. *2023 IEEE 9th Information Technology International Seminar (ITIS)*, 1–7. <https://doi.org/10.1109/ITIS59651.2023.10419994>
- Kedas, S., Kumar, A., & Jain, P. K. (2022). Dealing With Class Imbalance In Sentiment Analysis Using Deep Learning And SMOTE. In *Advances In Data Computing, Communication And Security* (Lecture No, Vol. 106, Pp. 407–416). Springer. https://doi.org/10.1007/978-981-16-8403-6_37
- Liang, H., Sun, X., Sun, Y., & Gao, Y. (2023). Comparison Of Text Preprocessing Methods. *Natural Language Engineering*, 29(3), 509–545. <https://doi.org/10.1017/S1351324922000213>
- Lin, C.-H., & Nuha, U. (2023). Sentiment Analysis Of Indonesian Datasets Based On A Hybrid Deep-Learning Strategy. *Journal Of Big Data*, 10(1), 88. <https://doi.org/10.1186/S40537-023-00782-9>
- Maulana, F. I., Rahayu, A., Widartha, V. P., Adi, P. D. P., Purnomo, A., & Pramono, A. (2024). A Comparative Study Of Machine Learning Models For Sentiment Analysis On User Reviews Of The Desa Wisata Nusantara Application In The Google Play Store. *2024 International Conference On ICT For Smart Society (ICISS)*, 1–7. <https://doi.org/10.1109/ICISS62896.2024.10751661>
- Purohit, S., Rajput, A., Vats, S., Mudgal, R., Shah, O. A., & Verma, A. (2024). Comparative Analysis Of LSTM And Random Forest Algorithms For Sentiment Classification In Movie Reviews. *2024 3rd International Conference On Applied Artificial Intelligence And Computing (ICAAIC)*, 1053–1057. <https://doi.org/10.1109/ICAAIC60222.2024.10575246>
- Samanmali, P. H. C., & Rupasingha, R. A. H. . (2024). Sentiment Analysis On Google Play Store App Users Reviews Based On Deep Learning Approach. *Multimedia Tools And Applications*, 83(20), 84425–84453. <https://doi.org/10.1007/S11042-024-19185-W>
- Santhosh Kumar, B., Praveen Yadav, P., & Panchala Prasad, P. (2025). Performance Analysis Of Machine Learning Algorithms On Imbalanced Datasets Using SMOTE Technique. In A. Kumar, V. K. Gunjan, S. Senatore, & Y.-C. Hu (Eds.), *Proceedings Of The 5th International Conference On Data Science, Machine Learning And Applications; Volume 1. ICDSMLA*

- 2023 (Pp. 147–156). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-8031-0_15
- Saputra, K. (2023). Multilabel Multiclass Sentiment And Emotion Dataset From Indonesian Mobile Application Review. *Data In Brief*, 50, 109576. <https://doi.org/10.1016/j.dib.2023.109576>
- Semary, N. A., Ahmed, W., Amin, K., Plawiak, P., & Hammad, M. (2024). Enhancing Machine Learning-Based Sentiment Analysis Through Feature Extraction Techniques. *PLOS ONE*, 19(2), 1–19. <https://doi.org/10.1371/journal.pone.0294968>
- Setiawan, B. (2024). A Review Of Sentiment Analysis Applications In Indonesia Between 2023-2024. *Journal Of Information Engineering And Educational Technology (JIEET)*, 8(2), 71–83. <https://doi.org/10.26740/jieet.v8n2.p71-83>
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment Analysis And Opinion Mining On Educational Data: A Survey. *Natural Language Processing Journal*, 2, 100003. <https://doi.org/10.1016/j.nlp.2022.100003>
- Sharma, A., & Ghose, U. (2021). Lexicon A Linguistic Approach For Sentiment Classification. *2021 11th International Conference On Cloud Computing, Data Science & Engineering (Confluence)*, 887–893. <https://doi.org/10.1109/Confluence51648.2021.9377057>
- Sun, Y., Li, Y., Zeng, Q., & Bian, Y. (2020). Application Research Of Text Classification Based On Random Forest Algorithm. *2020 3rd International Conference On Advanced Electronic Materials, Computers And Software Engineering (AEMCSE)*, 370–374. <https://doi.org/10.1109/AEMCSE50948.2020.00086>
- Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). Sentiment Analysis Of E-Government Applications On Google Play Using The Naive Bayes Algorithm. *JATISI (Jurnal Teknologi Informasi Dan Ilmu Komputer)*, 9(2), 785–795. <https://doi.org/10.35957/jatisi.v9i2.1835>
- Verma, S. (2022). Sentiment Analysis Of Public Services For Smart Society: Literature Review And Future Research Directions. *Government Information Quarterly*, 39(3), 101708. <https://doi.org/10.1016/j.giq.2022.101708>
- Wajidi, M. F., Inan, D. I., Juita, R., & Sanglise, M. (2024). STUDY ON THE QUALITY OF SERVICE OF THE MOBILE-BASED JKN APPLICATION: A SENTIMENT ANALYSIS APPROACH. *JIPi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(3), 1506–1517. <https://doi.org/10.29100/jipi.v9i3.5757> *
- Xu, Q. A., Chang, V., & Jayne, C. (2023). A Systematic Review Of Social Media-Based Sentiment Analysis: Emerging Trends And Challenges. *Decision Analytics Journal*, 6, 100172. <https://doi.org/10.1016/j.dajour.2022.100172>