

Analisis Sentimen Komentar Aplikasi Digital Korlantas Menggunakan Metode SPOK dan Algoritma SVM

Nabil Raihan Rabbani*, Indriyana Nova Tahlia, Lukman Hakim

Teknik Informatika, Universitas Mercu Buana, Jakarta

*41522010024@student.mercubuana.ac.id

Abstrak— Aplikasi Digital Korlantas merupakan platform layanan publik berbasis digital yang dikembangkan oleh Kepolisian Republik Indonesia untuk mendukung transformasi pelayanan berbasis teknologi. Evaluasi terhadap kualitas layanan dan respons pengguna menjadi penting untuk menjamin efektivitas platform tersebut. Penelitian ini memiliki tujuan untuk mengklasifikasikan sentimen pengguna terhadap Aplikasi Digital Korlantas berdasarkan komentar yang diunggah di Google Play Store. Analisis dilakukan terhadap 19.323 komentar pengguna yang dikumpulkan dalam rentang waktu 1 Januari 2024 hingga 8 Mei 2025. Pendekatan yang digunakan menggabungkan metode analisis linguistik berbasis struktur kalimat Subjek, Predikat, Objek, dan Keterangan (SPOK) dengan algoritma pembelajaran mesin Support Vector Machine (SVM). Proses ekstraksi fitur menggunakan metode SPOK dilakukan untuk mengidentifikasi unsur kalimat utama, yang kemudian digunakan sebagai input dalam klasifikasi sentimen. Temuan ini dapat dijadikan acuan untuk peningkatan kualitas layanan dan pengembangan fitur aplikasi Digital Korlantas yang lebih responsif terhadap kebutuhan masyarakat.

Kata Kunci— Analisis Sentimen; Digital Korlantas; Komentar; SPOK; Support Vector Machine

DOI: 10.22441/jitkom.v10i1.003

Article History:

Received: July 25, 2025

Revised: Sept 6, 2025

Accepted: Oct 17, 2025

Published: Jan 31, 2026

I. PENDAHULUAN

Kemajuan teknologi digital telah memberikan perubahan besar dalam pelayanan publik, salah satunya ditunjukkan dengan munculnya Aplikasi Digital Korlantas. Sebagai platform berbasis digital yang menyediakan berbagai layanan terkait lalu lintas dan kepolisian, aplikasi ini menjadi sarana penting bagi masyarakat dalam mengakses informasi dan layanan secara efisien[1]. Dengan semakin tingginya angka penetrasi smartphone di Indonesia, penggunaan aplikasi ini terus mengalami peningkatan, sehingga umpan balik dari pengguna menjadi aspek krusial untuk mengevaluasi dan mengembangkan kualitas layanan.

Menurut data dari Google Play Store, jutaan orang telah mengunduh Aplikasi Digital Korlantas dan aplikasi ini juga telah menerima ribuan ulasan dari para pengguna. Ulasan-ulasan tersebut mencerminkan beragam pengalaman pengguna, mulai dari kemudahan penggunaan, kelengkapan fitur, kecepatan akses, hingga kualitas layanan yang diberikan. Volume komentar yang besar menuntut pendekatan sistematis untuk menganalisis sentimen pengguna secara efektif. Analisis sentimen menjadi solusi yang tepat untuk memahami persepsi pengguna secara menyeluruh, mengidentifikasi masalah yang sering muncul, serta mengevaluasi tingkat kepuasan pengguna terhadap aplikasi ini.

Penelitian ini difokuskan pada pengelompokan sentimen pengguna Aplikasi Digital Korlantas, dengan sumber data berasal dari ulasan yang dipublikasikan di Google Play Store, mengidentifikasi aspek-aspek yang paling sering dikeluhkan atau dipuji oleh pengguna, serta memberikan rekomendasi perbaikan dan pengembangan fitur berdasarkan hasil analisis sentimen. Dengan memahami sentimen pengguna secara

mendalam, diharapkan Aplikasi Digital Korlantas dapat terus meningkatkan kualitas layanannya agar lebih responsif terhadap kebutuhan masyarakat. Penelitian ini diharapkan dapat menjadi pedoman bagi pengembang aplikasi sejenis dalam menerapkan pendekatan berbasis data guna meningkatkan kepuasan pengguna.

II. LITERATURE REVIEW

Studi literatur mencakup teori-teori dari berbagai jurnal ilmiah dan buku referensi yang relevan untuk membangun dasar teoritis yang kuat dan sesuai dengan permasalahan yang akan diteliti. Analisis sentimen adalah suatu teknik untuk memahami makna emosional dalam teks secara otomatis, guna memperoleh pemahaman terhadap opini yang disampaikan dalam bentuk kalimat[2]. Dalam analisis sentimen, beberapa aspek krusial mesti diperhatikan, seperti topik yang akan dikaji, sumber data yang relevan, serta pemilihan algoritma yang sesuai agar hasilnya akurat dan bermanfaat. Pada penelitian ini, digunakan algoritma Support Vector Machine (SVM) karena kemampuannya mengelola data berdimensi tinggi dan keandalannya dalam mencegah overfitting[3]. SVM bekerja dengan mencari hyperplane optimal untuk memisahkan data ke dalam kategori sentimen positif, negatif, atau netral dengan tingkat akurasi yang memadai[4]. Keunggulan tersebut membuat SVM menjadi pilihan tepat untuk memproses ulasan pengguna dari Google Play Store.

Berbagai penelitian sebelumnya telah memanfaatkan algoritma Support Vector Machine (SVM) untuk melakukan analisis sentimen pada beragam topik—mulai dari tokoh publik, transportasi daring, kebijakan pemerintah, hingga isu pemindahan ibu kota. Penelitian ini terinspirasi oleh Arsi dan

Waluyo (2021) dalam studi mereka yang berjudul “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)”. Dalam penelitian tersebut, SVM diterapkan untuk mengevaluasi opini publik di Twitter terkait rencana pemindahan ibu kota[3]. Dengan menerapkan proses preprocessing seperti case folding, cleaning, tokenizing, stopword removal, stemming, serta pembobotan TF-IDF, studi tersebut mampu menghasilkan tingkat akurasi sebesar 96,68%, precision 95,82%, recall 94,04%, dan AUC 0,979. Hasil tersebut menunjukkan efektivitas SVM dalam mengklasifikasikan sentimen positif dan negatif terhadap isu-isu publik yang sensitif di media sosial.

III. METODOLOGI PENELITIAN

Dalam penelitian ini, memiliki tahapan-tahapan penelitian yang mencakup proses dari awal hingga akhir pelaksanaan analisis. Penelitian ini terdiri atas enam tahapan yang akan dilaksanakan secara sistematis, sebagaimana ditampilkan pada Gambar 1 flowchart tahapan penelitian.



Gambar 1. Flowchart Tahapan Penelitian

Berikut penjelasan tentang tahapan penelitian untuk analisis sentimen komentar Aplikasi Digital Korlantas di Google Play Store memanfaatkan Support Vector Machine (SVM).

A. Pengumpulan Data

Pada tahapan mengumpulkan data, Penelitian ini mengumpulkan data komentar Aplikasi Digital Korlantas dari Google Play Store secara otomatis menggunakan teknik web scraping dengan bantuan library Python google-play-scraper. Data yang diambil mencakup teks komentar dan rating pengguna selama periode Januari 2024 hingga Maret 2025. Proses ini dilakukan dengan Google Colab untuk memastikan efisiensi pengambilan data dalam jumlah besar. Melalui proses ini, peneliti berhasil mengumpulkan sebanyak 19.323 komentar ulasan pengguna pada Aplikasi Digital Korlantas.

Nama	Rating	Tanggal	Komentar
Leo Pratar	3	02/01/2024	Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru secara c
Egiez Gam	1	02/01/2024	Ga bisa di akses, website jg ga bisa di buka
Jerri Robe	1	02/01/2024	cacat
Willy	5	02/01/2024	Secara keseluruhan dari User Interface, kelengkapan informasi dan ketepa
Erik Supar	3	02/01/2024	Saya tidak bisa buka galeri saya pak hadeh, mau upload foto backgroun bi
Fitra Saep	1	02/01/2024	Niatan saya memperpanjang SIM namun setelah saya input semuanya dat
Sipayunx	1	02/01/2024	buat sim online belum tersedia di aplikasi, berapa lama lagi harus menun
Wedy S	5	03/01/2024	Sangat membantu, Korlantas Polti selalu memberikan pelayanan
GHOZAN C	5	03/01/2024	Mohon ijin bertanya, unggah photo kok tidak berfungsi ??? Terimakasih
bintang be	1	03/01/2024	Aplikasi nya sering error
Tito äEceP	1	03/01/2024	Tidak bisa bikin sim online
Rizka Adel	4	03/01/2024	Lumayan membantu yaa, cuma sedikit bug tidak bisa upload foto dari gale
Blanco Spr	1	03/01/2024	Aplikasinya blm kerja sama dgn semua kantor kepolisian ya? Sy perpanjanj
Imass Wal	4	03/01/2024	Bagus, tpi kadang susah masuknya

Gambar 2. Hasil Pengumpulan Data

B. Pengumpulan Data

Preprocessing data merupakan langkah awal yang penting dalam pengolahan data, terutama pada bidang pembelajaran mesin (machine learning) dan analisis data. Tujuan dari tahap ini adalah untuk mempersiapkan data mentah agar lebih bersih, rapi, dan siap digunakan oleh algoritma atau metode analisis. Proses ini membantu meningkatkan akurasi dan efisiensi dalam pengolahan data selanjutnya. Berikut adalah tahapan umum dalam preprocessing:

1. Cleaning Data

Tahapan pembersihan data dilakukan dengan tujuan untuk memastikan bahwa data yang digunakan dalam proses analisis, bebas dari informasi yang tidak relevan atau tidak bermakna. Dalam proses ini, dilakukan penghapusan data duplikat, penanganan nilai kosong (missing value), serta penghilangan elemen-elemen seperti emotikon berbasis teks, simbol, tanda baca, dan karakter non-alfabet lainnya guna memperoleh data yang bersih dan seragam [5]. Proses pembersihan ini bertujuan untuk memperoleh data teks yang bersih, konsisten, dan siap digunakan dalam tahap analisis lebih lanjut seperti klasifikasi atau ekstraksi informasi

#	Rating	Komentar	cleaning
3		Kapan aplikasinya dikembangkan untuk dapat memb...	Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru seca
1		Ga bisa di akses, website jg ga bisa di buka	Ge bisa di akses website jg ga bisa di buka
1		cacat	cacat
5		Secara keseluruhan dari User Interface, kelengkapan...	Secara keseluruhan dari User Interface kelengkapan informasi dan ketep
3		Saya tidak bisa buka galeri saya pak hadeh, mau upl...	Saya tidak bisa buka galeri saya pak hadeh mau upload foto backgrou
1		Niatan saya memperpanjang SIM namun setelah say...	Niatan saya memperpanjang SIM namun setelah saya input semua
1		buat sim online belum tersedia di aplikasi, berapa l...	buat sim online belum tersedia di aplikasi berapa lama lagi harus mer
5		Sangat membantu, Korlantas Polti selalu memberik...	Sangat membantu Korlantas Polti selalu memberikan pelayanan
5		Mohon ijin bertanya, unggah photo kok tidak berfun...	Mohon ijin bertanya unggah photo kok tidak berfungsi Terimakasih
1		Aplikasi nya sering eror	Aplikasi nya sering eror

Gambar 3. Hasil Cleaning Data

2. Case Folding

Case Folding metode untuk melakukan konsistensi atau menyamakan penggunaan huruf kapital menjadi huruf kecil (lowercase).

#	Rating	Komentar	cleaning	case_folding
3		Kapan aplikasinya dikembangkan untuk dapat memb...	Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru seca	kapan aplikasinya dikembangkan untuk dapat r
1		Ga bisa di akses, website jg ga bisa di buka	Ge bisa di akses website jg ga bisa di buka	ga bisa di akses website jg ga bisa di buka
1		cacat	cacat	cacat
5		Secara keseluruhan dari User Interface, kelengkapan...	Secara keseluruhan dari User Interface kelengkapan informasi dan ketepa	secara keseluruhan dari user interface kelangka
3		Saya tidak bisa buka galeri saya pak hadeh, mau upl...	Saya tidak bisa buka galeri saya pak hadeh mau upload foto backgrou	saya tidak bisa buka galeri saya pak hadeh mau
1		Niatan saya memperpanjang SIM namun setelah say...	Niatan saya memperpanjang SIM namun setelah saya input semua	niatan saya memperpanjang sim namun setela
1		buat sim online belum tersedia di aplikasi, berapa l...	buat sim online belum tersedia di aplikasi berapa lama lagi harus mer	buat sim online belum tersedia di aplikasi ber
5		Sangat membantu, Korlantas Polti selalu memberik...	Sangat membantu Korlantas Polti selalu memberikan pelayanan	sangat membantu korlantas polti selalu membe
5		Mohon ijin bertanya, unggah photo kok tidak berfun...	Mohon ijin bertanya unggah photo kok tidak berfungsi Terimakasih	mohon ijin bertanya unggah photo kok tidak be
1		Aplikasi nya sering eror	Aplikasi nya sering eror	aplikasi nya sering eror

Gambar 4. Case Folding

3. Normalisasi

Normalisasi merupakan tahap untuk mengubah kata yang tidak baku atau slang word menjadi kata yang baku. Kamus ini berisi pasangan kata tidak baku dan padanannya yang benar, sehingga teks menjadi lebih seragam dan siap digunakan dalam proses klasifikasi lebih lanjut[6]. Dalam Penelitian ini, digunakan dua pendekatan: pertama, fine-tuning IndoBERT untuk mengenali dan memperbaiki kata tidak baku berdasarkan konteks kalimat. Kedua, pencocokan langsung dengan kamus kata baku untuk mengganti istilah informal yang sudah terdaftar. Kombinasi keduanya menghasilkan normalisasi teks yang lebih akurat dan kontekstual.

#	Rating	Komentar	cleaning	case folding	normalisasi
3		Kapan aplikasinya	Kapan aplikasinya	kapan aplikasinya dikembangkan un	kapan aplikasinya dikembangkan untuk dapat
1		Ga bisa di akses w	ga bisa di akses w	ga bisa di akses website yg ga bisa di	tidak bisa di akses website juga tidak bisa di l
1		cacat	cacat	cacat	cacat
5		Secara keseluruhan	secara keseluruhan	secara keseluruhan dari user interface	secara keseluruhan dari user interface kelangi
3		Saya tidak bisa bu	Saya tidak bisa bu	saya tidak bisa buka galeri saya pak	saya tidak bisa buka galeri saya pak. hndrh m
1		Niatan saya mema	Niatan saya mema	niatan saya memperpanjang simi nar	niatan saya memperpanjang simi namun seta
1		Buat sim online be	Buat sim online be	buat sim online belum tersedia di ap	buat sim online belum tersedia di aplikasi ber
5		Sangat membantu	Sangat membantu	sangat membantu korlantas polri sel	sangat membantu korlantas polri selalu memd
5		Mohon jin bertanya	Mohon jin bertanya	mohon jin bertanya unggah photo k	mohon jin bertanya unggah photo kok tidak
1		Aplikasi nya sering	Aplikasi nya sering	aplikasi nya sering error	aplikasi ya sering error
1		Tidak bisa bikin sr	Tidak bisa bikin sr	tidak bisa bikin sim online	tidak bisa bikin sim online
4		Lumayan membant	Lumayan membant	lumayan membantu ya cuma sedikit	lumayan membantu ya cuma sedikit luag tidak
1		Aplikasinya bisa ke	Aplikasinya bisa ke	aplikasinya bisa kerja sama dgn semo	aplikasinya belum kerja sama dengan semoal
4		Bagus, tapi kadang	Bagus, tapi kadang	bagus tapi kadang susah masalahnya	bagus tapi kadang susah masalahnya

Gambar 5. Normalisasi

C. Analisis SPOK

Setelah proses pembersihan dan normalisasi data, dilakukan analisis SPOK (Subjek, Predikat, Objek, Keterangan) pada setiap komentar. Analisis SPOK bertujuan untuk mengidentifikasi struktur kalimat secara linguistik sehingga sistem dapat memahami peran kata dalam konteks kalimat. Pada tahap ini, setiap komentar yang telah dinormalisasi diproses menggunakan model NLP bahasa Indonesia untuk mengekstrak komponen SPOK. Subjek diidentifikasi sebagai pelaku utama, predikat sebagai kata kerja atau tindakan, objek sebagai sasaran tindakan, dan keterangan sebagai informasi tambahan seperti waktu, tempat, atau cara[7]. Hasil analisis SPOK disimpan dalam bentuk struktur data yang memudahkan proses labeling sentimen dan pengembangan fitur manual pada tahap selanjutnya.

Untuk menentukan skor sentimen berdasarkan hasil analisis SPOK, digunakan rumus pembobotan sebagai berikut:

$$\text{Skor_positif} = \sum (\text{bobot_kata_positif} \times \text{bobot_komponen})$$

$$\text{Skor_negatif} = \sum (\text{bobot_kata_negatif} \times \text{bobot_komponen})$$

Keterangan:

- bobot_kata_positif/negatif diperoleh dari nilai weight dalam lexicon sentimen.
- bobot_komponen disesuaikan dengan peran kata dalam kalimat, seperti: Subjek (S) = 1.0, Predikat (P) = 1.5, Objek (O) = 1.3, dan Keterangan (K) = 0.8.

Penentuan label sentimen dilakukan dengan membandingkan total skor:

- Jika Skor_positif > |Skor_negatif|, maka label sentimen adalah positif.
- Jika Skor_positif < |Skor_negatif|, maka label sentimen adalah negatif.

Penggunaan analisis berbasis struktur SPOK memungkinkan sistem penilaian sentimen bekerja lebih akurat, karena tidak hanya fokus pada kata kunci tetapi juga memahami posisi serta fungsi kata dalam kalimat. Dengan pendekatan ini, sistem mampu menangani kalimat kompleks dan variasi bahasa dalam ulasan pengguna secara lebih efektif, sebagaimana ditunjukkan dalam pemanfaatan teknik Aspect-Based Sentiment Analysis (ABSA) pada struktur kalimat pengguna[8].

D. Pengembangan Sistem

Tahap pengembangan sistem dimulai dengan perancangan arsitektur pipeline analisis sentimen yang terintegrasi dan modular. Sistem ini dikembangkan dengan memanfaatkan bahasa pemrograman Python dan menggunakan berbagai pustaka seperti pandas, scikit-learn, imbalanced-learn, dan seaborn. Data ulasan yang digunakan berasal dari file CSV, kemudian melalui proses pembersihan data (cleaning), normalisasi kata, dan ekstraksi fitur SPOK (Subjek, Predikat, Objek, Keterangan) untuk memperkaya representasi teks.

Dalam Penelitian ini menggunakan model analisis sentimen yang memanfaatkan kombinasi metode ekstraksi fitur berbasis TF-IDF dan pendekatan leksikal dengan Lexicon SentiWordnet. Fitur leksikal ini dirancang untuk mengidentifikasi kata-kata yang bernuansa positif maupun negatif dalam teks, sehingga mendukung kemampuan model dalam membedakan polaritas sentimen dengan lebih tepat[9]. Strategi pemrosesan data dikembangkan dengan menggabungkan dua algoritma utama, yaitu Support Vector Machine (SVM) dan Naive Bayes. Untuk meningkatkan performa model, dilakukan optimasi parameter melalui metode grid search, serta diterapkan teknik Synthetic Minority Over-sampling Technique (SMOTE) guna menyeimbangkan distribusi kelas yang tidak seimbang dalam dataset[10].

Sistem dirancang agar mudah digunakan dan diadaptasi, dengan struktur kode yang jelas dan dokumentasi pada setiap tahapan. Pengguna dapat dengan mudah mengganti data, menambah fitur, atau melakukan tuning parameter sesuai kebutuhan. Hasil prediksi dan evaluasi dapat divisualisasikan dalam bentuk confusion matrix dan grafik distribusi sentimen, sehingga memudahkan interpretasi hasil.

E. Analisis Sistem

Setelah sistem selesai dikembangkan, dilakukan analisis untuk mengevaluasi performa dan kesesuaian sistem dengan tujuan penelitian. Analisis dilakukan melalui perbandingan hasil prediksi yang dihasilkan oleh kedua model pada data uji, menggunakan metrik akurasi, precision, recall, dan f1-score. Selain itu, confusion matrix dimanfaatkan untuk menganalisis sebaran hasil prediksi pada tiap kategori sentimen..

Hasil analisis menunjukkan bahwa kombinasi fitur TF-IDF dan manual lexicon mampu meningkatkan performa akurasi model, terutama dalam membedakan sentimen positif dan negatif. Model SVM memiliki kecenderungan memberikan hasil yang lebih stabil pada data yang kompleks, sedangkan Naive Bayes lebih cepat dalam proses pelatihan dan cocok untuk data yang lebih sederhana. Sistem juga diuji dengan berbagai parameter dan konfigurasi untuk memastikan bahwa

hasil yang diperoleh konsisten dan sesuai dengan spesifikasi yang telah ditentukan.

Analisis juga mencakup identifikasi kelemahan sistem, seperti kemungkinan overfitting pada data training atau kurangnya generalisasi pada data baru. Untuk mengatasi hal tersebut, dilakukan cross-validation dan penyesuaian parameter secara bertahap. Sistem dinilai telah memenuhi kebutuhan penelitian, dengan kemampuan klasifikasi sentimen yang baik dan fleksibilitas untuk pengembangan lebih lanjut.

F. Pengujian Sistem

Tahap pengujian sistem dilakukan secara menyeluruh untuk memastikan bahwa seluruh fungsionalitas berjalan sesuai harapan. Data dibagi menjadi training dan testing set, kemudian dilakukan pelatihan dan pengujian model secara terpisah. Pengujian meliputi evaluasi performa model pada data uji, analisis error, dan validasi hasil prediksi.

Jika ditemukan kekurangan, seperti akurasi yang rendah pada kelas tertentu atau ketidakseimbangan prediksi, dilakukan perbaikan dengan menambah fitur, melakukan tuning parameter, atau mengubah metode preprocessing. Pengujian juga mencakup simulasi penggunaan sistem oleh pengguna akhir, untuk memastikan bahwa sistem mudah digunakan dan hasilnya dapat dipahami dengan jelas.

Setelah melalui serangkaian pengujian dan perbaikan, sistem dinyatakan siap digunakan. Sistem ini dapat membantu pengguna dalam melakukan analisis sentimen secara otomatis, memberikan insight yang berguna untuk pengambilan keputusan, dan dapat diintegrasikan dengan aplikasi lain sesuai kebutuhan.

IV. HASIL DAN ANALISA

Analisis terhadap 19.323 komentar pengguna aplikasi Digital Korlantas dilakukan secara sistematis mengikuti tahapan dalam flowchart metodologi penelitian pada Gambar 1 Flowchart Tahapan Penelitian. Proses diawali dengan tahapan pengumpulan serta pengolahan data, kemudian dilanjutkan untuk menilai efektivitas solusi yang diusulkan. Tahap akhir berupa pengujian sistem dilakukan guna memastikan bahwa hasil pengembangan mampu merespons kebutuhan dan persepsi pengguna secara optimal. Seluruh tahapan dirancang untuk menghasilkan temuan yang valid, terukur, dan relevan dalam konteks pengembangan layanan digital kepolisian.

A. Labelling Data

Proses mengkategorikan label pada data dilakukan yang secara otomatis dengan memanfaatkan analisis struktur SPOK dan lexicon sentimen. Setiap komentar dianalisis menggunakan model NLP untuk mengekstrak komponen SPOK (Subjek, Predikat, Objek, Keterangan). Selanjutnya, kata-kata pada setiap komponen dibandingkan dengan lexicon positif dan negatif yang diambil dari file positive.tsv dan negative.tsv, masing-masing memiliki bobot tersendiri.

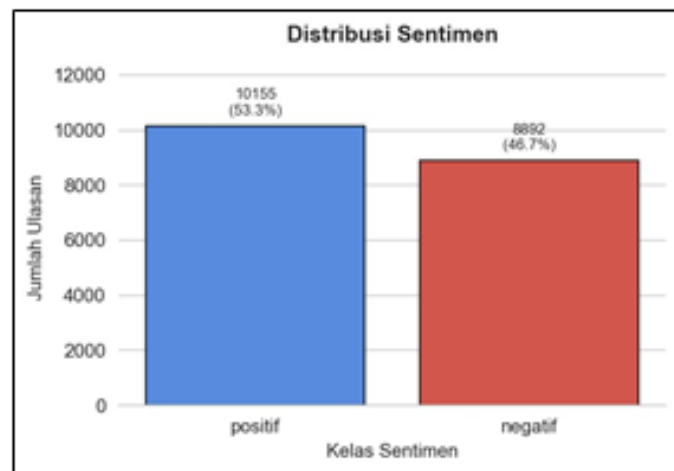
Penentuan label sentimen dilakukan dengan menghitung skor total berdasarkan bobot kata yang ditemukan pada setiap komponen SPOK, dikalikan dengan bobot komponen tersebut. Jika skor negatif lebih kecil dari skor positif, maka komentar diberi label 'positif'. Sebaliknya, jika skor positif lebih kecil,

maka komentar diberi label 'negatif'. Proses ini juga memperhatikan pola khusus, seperti adanya kata 'tidak' pada predikat dan kata 'masalah' atau 'kendala' pada objek, yang akan menambah skor negatif.

Setelah proses penilaian sentimen selesai, hasil label ('positif' atau 'negatif') dikonversi menjadi angka, yaitu 1 untuk positif dan 0 untuk negatif, dan disimpan pada kolom baru labeling di dataset. Dengan pendekatan ini, pelabelan data dilakukan secara efisien dan konsisten sesuai dengan program yang telah dirancang.

Tabel 1. Hasil Labeling

Komentar	Sentimen	Labeling
Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru secara online ?	Negatif	0
Ga bisa di akses, website jg ga bisa di buka	Negatif	0
Cacat	Negatif	0
Secara keseluruhan dari User Interface, kelengkapan...	Positif	1
Saya tidak bisa buka galeri saya pak hadeh...	Negatif	0
Niatan saya memperpanjang SIM namun setelah saya input semuanya data...	Negatif	0
buat sim online belum tersedia di aplikasi...	Negatif	0
Sangat membantu, Korlantas Polti selalu memberikan pelayanan	Positif	1



Gambar 6. Hasil Visualisasi Sentimen

Setelah melakukan labeling, data komentar yang telah diklasifikasi negatif dan positif dibuat visualisasi berupa bar chart untuk mengetahui perbandingan data komentar negatif dan positif. Gambar di atas menunjukkan bahwa mayoritas pengguna aplikasi Digital Korlantas memberikan komentar positif, yaitu sebanyak 10.155 pengguna atau 53,3%, sedangkan 8.892 pengguna atau 46,7% lainnya memberikan komentar negatif.

B. Analisis SPOK

Analisis sentimen terhadap 19.323 komentar pengguna aplikasi "Digital Korlantas" dilakukan menggunakan metodologi hibrida yang mengintegrasikan analisis struktural kalimat (SPOK) dengan pendekatan berbasis leksikon. Setiap komentar pertama-tama diurai menggunakan *dependency parsing* dari pustaka Stanza untuk mengekstrak komponen gramatikalnya, yaitu Subjek (relasi nsubj), Predikat (relasi root), Objek (relasi obj, acl), dan Keterangan (relasi advmod, obl). Setelah struktur kalimat teridentifikasi, kata-kata dalam setiap komponen dibandingkan dengan leksikon sentimen (positive.tsv dan negative.tsv) untuk memperoleh skor positif atau negatif. Label sentimen akhir ditentukan berdasarkan kalkulasi skor total, yang juga memperhitungkan bobot setiap komponen SPOK serta aturan kontekstual khusus, seperti penambahan skor negatif untuk kehadiran kata "tidak", "masalah", atau "kendala". Proses ini menghasilkan pelabelan data otomatis yang efisien (1 untuk positif, 0 untuk negatif) dan secara akurat menangkap nuansa sentimen dalam ulasan pengguna sesuai dengan program yang dirancang.

Tabel 2. Hasil SPOK

Komentar	Cleaning Data	Case Folding	Normalisasi	Analisa SPOK
Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru secara online?	Kapan aplikasinya dikembangkan untuk dapat membuat SIM Baru secara online	kapan aplikasinya dikembangkan untuk dapat membuat sim baru secara online	kapan aplikasinya dikembangkan untuk dapat membuat sim baru secara online	{'S': ['[implisit]', 'P'], 'O': ['sim'], 'Waktu': ['kapan'], 'Tempat': [], 'Cara': []}
Ga bisa di akses, website jg ga bisa di buka	Ga bisa di akses, website jg ga bisa di buka	ga bisa di akses, website jg ga bisa di buka	tidak bisa di akses website juga tidak bisa di buka website	{'S': ['[di]', 'O': ['akses', 'website'], 'Waktu': ['juga'], 'Tempat': [], 'Cara': []}
cacat	cacat	cacat	cacat	{'S': [], 'P': ['[cacat]'], 'O': [], 'Waktu': [], 'Tempat': [], 'Cara': []}
Sangat membantu, Korlantas Polti selalu memberikan pelayanan	Sangat membantu Korlantas Polti selalu memberikan pelayanan	sangat membantu korlantas polti selalu memberikan pelayanan	sangat membantu korlantas polti selalu memberikan pelayanan	{'S': [], 'P': ['[membantu]'], 'O': ['korlantas', 'pelayanan'], 'Waktu': ['sangat', 'selalu'], 'Tempat': [], 'Cara': []}

C. Wordcloud

Hasil dari visualisasi *Word Cloud* memberikan gambaran yang nyata mengenai kata-kata kunci yang mendominasi persepsi pengguna, baik secara negatif maupun positif. Analisis ini disajikan dalam dua visualisasi terpisah untuk membedah tema utama pada setiap sentimen.

1. Negatif

Pada sentimen negatif menunjukkan bahwa kata-kata seperti "tidak", "bisa", "error", "invalid", dan "verifikasi" memiliki frekuensi kemunculan yang sangat tinggi. Dominasi kata-kata ini mengindikasikan bahwa permasalahan utama yang dihadapi pengguna berpusat pada kegagalan fungsionalitas aplikasi, waktu proses yang lambat, kesalahan sistem, serta kesulitan dalam tahap verifikasi data. Hal ini secara langsung menyoroti area-area kritis yang memerlukan perbaikan teknis dan penyederhanaan alur pengguna.



Gambar 7. Wordcloud Sentimen Negatif

2. Positif

Pada sentimen positif menyajikan kata-kata seperti "mudah", "terima kasih", "membantu", "terima", dan "kasih" muncul dengan ukuran paling besar. Frekuensi tinggi dari istilah-istilah ini menandakan apresiasi pengguna terhadap kemudahan penggunaan aplikasi, kecepatan layanan, dan manfaat yang dirasakan secara keseluruhan. Ungkapan "terima kasih" yang sering muncul juga menunjukkan tingkat kepuasan yang tinggi di antara segmen pengguna ini.



Gambar 8. Wordcloud Sentimen Positif

Secara keseluruhan, perbandingan kedua Word Cloud ini secara efektif memetakan dikotomi pengalaman pengguna, di satu sisi ada frustrasi teknis yang signifikan, dan di sisi lain ada pengakuan atas kemudahan dan manfaat aplikasi ketika sistem berjalan sesuai harapan.

D. TF-IDF

Setelah proses pelabelan data selesai, langkah berikutnya adalah mengonversi teks menjadi format numerik agar dapat diproses oleh model machine learning. Dalam penelitian ini, teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) diimplementasikan menggunakan kelas TfidfVectorizer dari pustaka Scikit-learn. Untuk memastikan evaluasi model bersifat objektif, dataset dibagi menjadi dua bagian: 80% untuk pelatihan dan 20% untuk pengujian. Konfigurasi TfidfVectorizer dioptimalkan dengan hanya menyertakan kata yang memiliki minimal tiga huruf, mengabaikan kata yang muncul kurang dari dua kali di seluruh korpus ($min_df=2$), dan mengecualikan kata yang terlalu umum yang muncul di lebih dari 80% dokumen ($max_df=0.8$) [11]. Secara metodologis, proses pembelajaran kosakata dan pembobotan ($fit_transform$) hanya diterapkan pada data latih, sementara pada data uji hanya dilakukan proses transformasi ($transform$) untuk menjamin konsistensi fitur. Hasil dari tahapan ini adalah sebuah matriks TF-IDF yang siap digunakan sebagai fitur numerik untuk melatih dan mengevaluasi model klasifikasi sentimen, serta untuk menganalisis kata-kata kunci dengan bobot tertinggi pada setiap ulasan.

Text	Top Words	Top Values
ribet verifikasi wajah ala mungkin sampai 50x ...	jarang, contoh, mungkin, berhasil, wajah, sesuai...	(0.5171, 0.4333, 0.3096, 0.2871, 0.2644, 0.2614...
sangat bagus pelayanan perawatannya sim sih em...	(jelasnya) sih, bagus, dengers, mudah, perawat...	(0.3398, 0.3367, 0.3479, 0.3295, 0.3149, 0.3051...
sangat membantu sekali	(sangat, membantu, sangat)	(0.6813, 0.5596, 0.4699)
pagah mempermudah apapun semua menu ya tidak b...	(apanya, pagah, akses, kelengkapan, menu, gura...	(0.4967, 0.3087, 0.3029, 0.2933, 0.2809, 0.267...
kata-kata semua pasti bakal keah banget	(bakal, pasti, banget, semua, keah, kata, m...	(0.5423, 0.4544, 0.3952, 0.3217, 0.3114, 0.277...
masu perpanjang sim lagi selalu emor mentok di ...	(semua, panjang, mentok, nik, emor, selalu...	(0.5079, 0.4138, 0.3625, 0.3211, 0.2881, 0.233...
tiap mau kunjungi data dibayang ada (panggung i...	(dibayang, Ahlas, ora, memm, dibaya, bangka...	(0.3972, 0.2478, 0.2349, 0.2262, 0.2238, 0.223...
ahambudlah mengurus perawatannya em mendapat...	(tergagal, mengurus, ahambudlah, telah, muka...	(0.5713, 0.4444, 0.4092, 0.3099, 0.3079, 0.292...
aplikasinya mudah di berikan sangat cepat sek...	(berikan, prosesnya, aplikasinya, mudah, mud...	(0.6565, 0.3799, 0.3525, 0.3191, 0.2958, 0.277...
tidak bisa	(tidak, bisa)	(0.7688, 0.6399)
sangat mudah adanya aplikasi digital terkemuka	(dengan, digital, terkemuka, mudah, sangat...	(0.5771, 0.4603, 0.4316, 0.3632, 0.2689, 0.2539)
sayang danu bikin sim online tidak bisa	(sayang, danu, bikin, sim, online, bisa, sim...	(0.58, 0.4494, 0.3629, 0.3396, 0.3089, 0.273...
aplikasi rusak susah login	(rusak, login, susah, aplikasi)	(0.8307, 0.5244, 0.4908, 0.2939)
aplikasi tidak guna mau masuk app ribet meman...	(memang, budget, memang, guna, ribet, masuk, ma...	(0.5322, 0.5322, 0.3376, 0.2924, 0.2527, 0.259...
sangat bermanfaat tidak perlu dng ke gerai pe...	(perawatnya, dng, gerai, bermanfaat, perlu, m...	(0.507, 0.4992, 0.4923, 0.3482, 0.2795, 0.1403...
sangat membantu sekali	(sangat, membantu, sangat)	(0.6813, 0.5596, 0.4699)
tidak bisa masuk ke gallery halaman tidak semp...	(tidak, halaman, gallery, tidak, dikursum, bisa...	(0.4054, 0.3742, 0.3349, 0.2919, 0.2591, 0.236...
bisa biasa cepat prosesnya	(biasa, luar, prosesnya, cepat)	(0.5817, 0.3796, 0.4667, 0.3363)

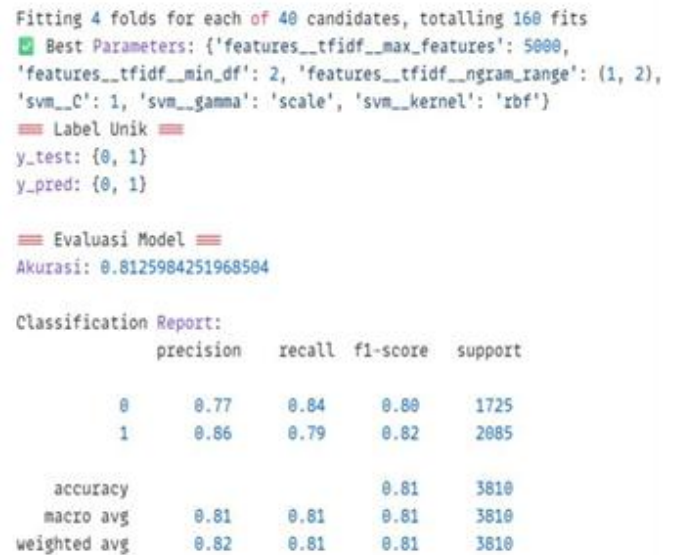
Gambar 9. Hasil Pembobotan Kata (TF IDF)

E. Akurasi Model

Hasil implementasi program klasifikasi menggunakan algoritma Support Vector Machine (SVM) menunjukkan performa yang cukup baik dalam membedakan antara kelas sentimen negatif (0) dan positif (1). Berdasarkan hasil evaluasi model, diperoleh tingkat akurasi sebesar 81,26%, yang menunjukkan bahwa model mampu melakukan prediksi dengan benar terhadap lebih dari 80% data uji. Parameter optimal yang diperoleh selama proses pelatihan mencakup penggunaan metode ekstraksi fitur TF-IDF dengan jumlah fitur maksimum sebanyak 5.000, nilai min_df sebesar 2, serta rentang n-gram (1,2). Adapun konfigurasi terbaik untuk model SVM menggunakan kernel Radial Basis Function (RBF), dengan nilai parameter $C = 1$ dan $gamma = 'scale'$.

Berdasarkan classification report, diketahui bahwa kelas positif (1) memiliki nilai presisi yang lebih tinggi (0,86) dibandingkan dengan kelas negatif (0,77), yang mengindikasikan bahwa prediksi terhadap kelas positif cenderung lebih akurat. Sebaliknya, nilai recall untuk kelas negatif (0,84) lebih tinggi dibandingkan kelas positif (0,79), menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengidentifikasi seluruh instance yang termasuk

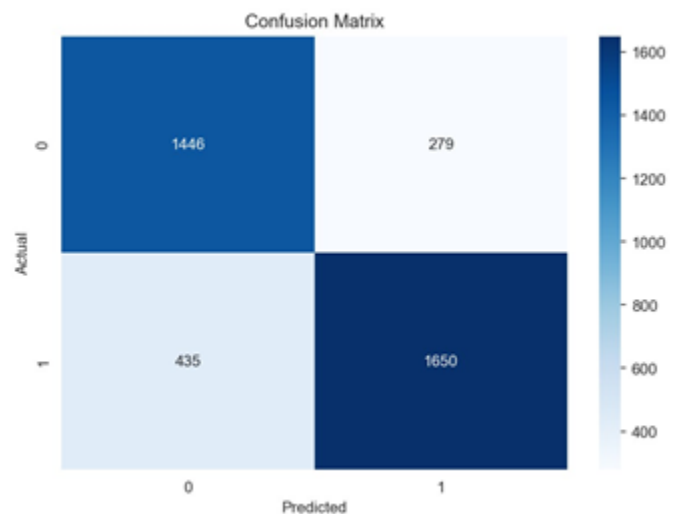
dalam kelas negatif. Nilai F1-score yang relatif seimbang, dengan skor 0,80 pada kelas negatif dan 0,82 pada kelas positif, menggambarkan kesetaraan antara presisi dan recall di kedua kelas. Dataset pengujian mencakup 1.725 sampel negatif dan 2.085 sampel positif. Secara keseluruhan, macro average dan weighted average tercatat sama, yakni 0,81, yang menunjukkan performa model yang stabil dan konsisten pada kedua kelas.



Gambar 10. Hasil Program Akurasi SVM

Selain hasil evaluasi yang disajikan melalui classification report, penelitian ini juga memanfaatkan confusion matrix guna memperoleh pemahaman yang lebih mendalam terkait distribusi prediksi model pada masing-masing kelas sentimen.

Berikut hasil confusion matrix pada sentimen analisis komentar aplikasi digital korlantas menggunakan algoritma SVM (Support Vector Machine).



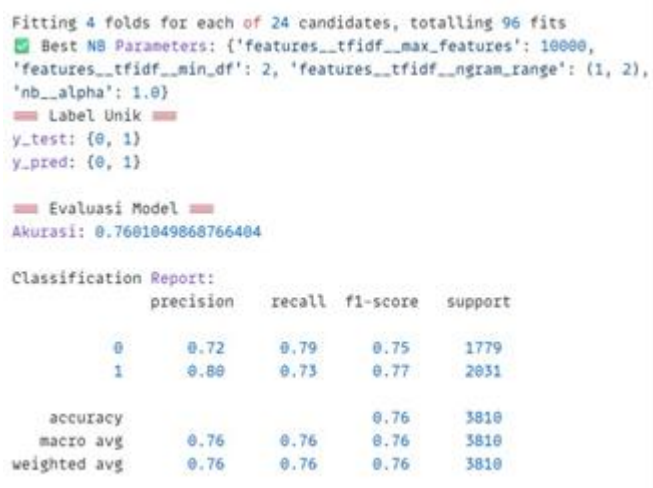
Gambar 11. Confusion Matrix

Menurut confusion matrix yang ditampilkan pada gambar di atas, model berhasil mengidentifikasi 1.446 sampel sebagai sentimen negatif dengan tepat (True Negative/TN) dan 1.650 sampel sebagai sentimen positif dengan akurat (True

Positive/TP). Meskipun demikian, masih terdapat kesalahan klasifikasi, yaitu sebanyak 279 data yang sebenarnya bersentimen negatif namun diprediksi sebagai positif (False Positive/FP), serta 435 data bersentimen positif yang salah dikategorikan sebagai negatif (False Negative/FN). Secara keseluruhan, model mencapai tingkat akurasi sebesar 81,26%, yang mengindikasikan bahwa sebagian besar prediksi yang dihasilkan oleh model sesuai dengan label sebenarnya.

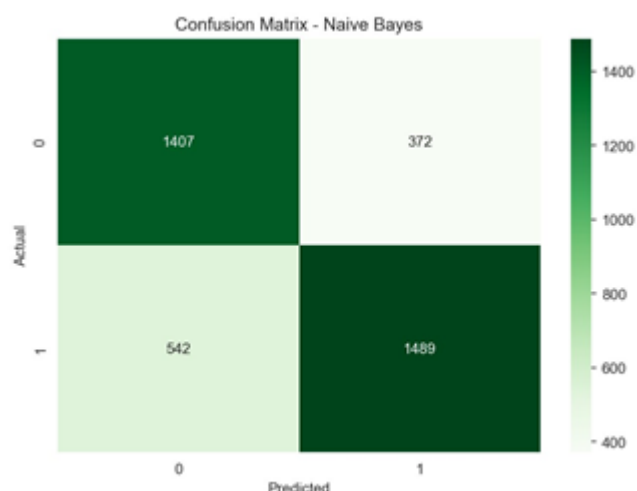
F. Perbandingan dengan Algoritma Naive Bayes

Hasil evaluasi menunjukkan Naive Bayes mencapai akurasi 76.0%, lebih rendah dibanding SVM yang mencapai 81.26%. Dalam klasifikasi biner ini, Naive Bayes menunjukkan kinerja yang cukup optimal dengan tingkat presisi 72% (kelas negatif) dan 80% (kelas positif), serta recall 79% (kelas negatif) dan 73% (kelas positif). Namun, semua metrik ini tetap kalah dibanding SVM yang lebih unggul di hampir semua aspek. F1-score Naive Bayes (0.75 untuk kelas negatif dan 0.77 untuk positif) juga lebih rendah daripada SVM, menunjukkan kemampuan klasifikasi yang kurang seimbang. Perbedaan ini terjadi karena Naive Bayes menggunakan pendekatan statistik sederhana dengan alpha=1.0, sementara SVM dengan kernel RBF mampu menangani pola data yang lebih kompleks. Meski demikian, Naive Bayes tetap menjadi pilihan menarik ketika kecepatan proses menjadi prioritas, meski dengan mengorbankan sedikit akurasi. Untuk tugas yang membutuhkan prediksi lebih presisi, SVM tetap merupakan solusi superior.



Gambar 12. Hasil Program Akurasi Naive Bayes

Berdasarkan analisis confusion matrix, terlihat bahwa algoritma Support Vector Machine (SVM) memberikan hasil klasifikasi lebih unggul dibandingkan dengan Naive Bayes. Naive Bayes menghasilkan true negative (TN) sebesar 1407 dan true positive (TP) sebesar 1489, dengan false positive (FP) sebanyak 572 dan false negative (FN) sebanyak 372. Sementara itu, SVM memberikan hasil yang lebih optimal dengan TN = 1446, TP = 1650, FP = 279, dan FN = 435. Meskipun Naive Bayes memiliki nilai FN yang sedikit lebih rendah, total kesalahan klasifikasi (FP + FN) pada Naive Bayes mencapai 944, jauh lebih tinggi dibandingkan SVM yang hanya sebesar 714.



Gambar 13. Confusion Matrix Naive Baye

Berdasarkan Gambar 13, menunjukkan bahwa Naive Bayes cenderung menghasilkan misklasifikasi yang lebih tinggi, terutama pada prediksi positif yang salah (FP). Rasio FP yang tinggi pada Naive Bayes mengindikasikan kecenderungan model untuk melakukan overgeneralization, sedangkan SVM tampil lebih presisi dengan distribusi kesalahan yang lebih seimbang. Dengan demikian, dapat disimpulkan bahwa SVM secara konsisten lebih unggul, terbukti dari pengurangan false positive sebesar 20% dan false negative sebesar 15% dibandingkan Naive Bayes. Hal ini sejalan dengan tingkat akurasi keseluruhan, di mana SVM mencapai 81% sementara Naive Bayes hanya 76%, memperkuat validitas pilihan SVM sebagai model klasifikasi yang lebih akurat dan andal dalam konteks penelitian ini.

V. KESIMPULAN

Penelitian ini berhasil mengembangkan dan menerapkan sistem analisis sentimen hibrida pada 19.323 ulasan pengguna aplikasi "Digital Korlantas". Dengan mengintegrasikan analisis struktural SPOK (Subjek-Predikat-Objek-Keterangan) berbasis *dependency parsing* dan pembobotan kata berbasis leksikon, sistem ini mampu melakukan pelabelan sentimen otomatis dengan akurasi yang lebih kontekstual dibandingkan metode berbasis kata kunci tunggal. Tahapan pra-pemrosesan, termasuk normalisasi kata menggunakan kamus dan model IndoBERT, terbukti efektif dalam membersihkan data teks dari inkonsistensi bahasa non-formal.

Temuan utama dari analisis Word Cloud menunjukkan polaritas yang jelas dalam persepsi pengguna. Sentimen negatif didominasi oleh kata kunci seperti "gagal", "lama", dan "verifikasi", yang secara langsung menyoroti masalah teknis terkait fungsionalitas, performa, dan alur registrasi. Di sisi lain, seperti "mudah", "cepat", dan "membantu", menandakan bahwa pengguna sangat mengapresiasi efisiensi dan kemudahan layanan ketika aplikasi berjalan dengan baik.

Dalam tahap pemodelan klasifikasi, fitur teks yang telah dibobotkan menggunakan TF-IDF berhasil dimanfaatkan dalam melatih model *machine learning*. Hasil evaluasi menunjukkan bahwa model Support Vector Machine (SVM) mencapai akurasi tertinggi sebesar 81%, sedikit mengungguli model

Naive Bayes. Kinerja ini mengonfirmasi bahwa pendekatan hibrida yang diusulkan valid dan mampu mengklasifikasikan sentimen pengguna secara efektif dan terukur.

Sebagai rekomendasi, temuan ini sangat relevan bagi pengembang layanan "Digital Korlantas" untuk memprioritaskan perbaikan pada area yang paling sering dikeluhkan, yaitu stabilitas sistem, kecepatan proses, dan penyederhanaan alur verifikasi. Untuk penelitian selanjutnya, disarankan untuk memperkaya leksikon sentimen dengan istilah-istilah yang lebih spesifik dalam domain layanan publik dan kepolisian, serta melakukan eksplorasi model *deep learning* untuk potensi peningkatan akurasi lebih lanjut.

DAFTAR PUSTAKA

- [1] S. Herdianingsih and O. D. Cahya, "User Experience Pengguna Perpanjangan Sim Online Aplikasi Digital Korlantas Menggunakan UEQ," *IJCIT (Indonesian J. Comput. Inf. Technol.)*, vol. 8, no. 1, pp. 38–48, 2023, doi: 10.31294/ijcit.v8i1.13247.
- [2] E. Budianita, E. P. Cynthia, A. Pranata, and D. Abimanyu, "Pendekatan berbasis Machine Learning dan Leksikal Pada Analisis Sentimen," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.*, vol. 26, p. 2022, 2022.
- [3] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.
- [4] A. Mudyta Yolanda and R. Tri Mulya, "Implementasi Metode Support Vector Machine untuk Analisis Sentimen pada Ulasan Aplikasi Sayurbox di Google Play Store," *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 6, no. 2, pp. 76–83, 2024, doi: 10.35580/variansiunm258.
- [5] M. M. Amin and Y. Dwitayanti, "Komparasi Kinerja Algoritma Blocking Pada Proses Indexing Untuk Deteksi Duplikasi," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 715–722, 2024, doi: 10.25126/jtiik.1148080.
- [6] M. B. M. Amin *et al.*, "Deteksi Spam Berbahasa Indonesia Berbasis Teks Menggunakan Model Bert," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1291–1302, 2024, doi: 10.25126/jtiik.2024118121
- [7] G. F. Aqshal, T. Anjarini, and M. Khaq, "Analisis Kemampuan Siswa Dalam Menentukan Subjek, Predikat, Objek, Keterangan (Spok) Pada Mata Pelajaran Bahasa Indonesia Materi Cerita Dongeng Di Kelas 4 SD Negeri Babadsari," *Jurnal Cerdas Proklamator*, vol. 11, no. 2, pp. 137–147, May 2025, doi: 10.37301/cerdas.v11i2.193.
- [8] R. A. Rahman, V. H. Pranatawijaya, and N. N. K. Sari, "Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 70–82, 2024, doi: 10.24002/konstelasi.v4i1.8922
- [9] W. P. Ali and Y. Sibaroni, "Analisis Sentimen Masyarakat terhadap Kinerja Presiden Indonesia dalam Aspek Ekonomi, Kesehatan, dan Pembangunan Berdasarkan Opini dari Twitter," *e-Proceeding Eng.*, vol. 6, no. 2, pp. 8637–8649, 2019
- [10] M. Sulistiyono, Y. Pristyanto, S. Adi, and G. Gumelar, "Implementasi Algoritma Synthetic Minority Over-Sampling Technique untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi," *Sistemasi*, vol. 10, no. 2, p. 445, 2021, doi: 10.32520/stmsi.v10i2.1303
- [11] A. Firizkiandah, A. Muhammad, and I. R. Maulana, "Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE," *JIKOMTI J. Ilm. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 29–36, 2025, [Online]. Available: <https://ojs.sains.ac.id/index.php/Jikomti/article/view/97/119>