

Analisis Sentimen Pada Aplikasi Identitas Kependudukan Digital Menggunakan Metode Support Vector Machine (SVM)

1st Dito Putra
Sistem Informasi
Telkom University
 Surabaya, Indonesia

ditoputra@student.telkomuniversity.ac.id

2nd Sri Hidayati
Sains Data
Telkom University
 Surabaya, Indonesia

srihidayatisi@telkomuniversity.ac.id

3rd Anisa Dzulkarnain
Sistem Informasi
Telkom University
 Surabaya, Indonesia

anisadzulkarnain@telkomuniversity.ac.id

Abstrak — Aplikasi Identitas Kependudukan Digital (IKD) merupakan gambaran digital dari dokumen kependudukan yang menampilkan data pribadi pengguna melalui perangkat elektronik. Rating aplikasi IKD di Play Store yang hanya 3,5 dari 5 menunjukkan tingkat kepuasan pengguna yang masih rendah. Tujuan dari penelitian ini adalah untuk menganalisis sentimen terhadap ulasan aplikasi IKD di Play Store. Data dikumpulkan melalui proses scraping, kemudian dilabeli secara manual menjadi dua kelas, yaitu positif dan negatif. Selanjutnya, dilakukan preprocessing dan pembobotan data menggunakan TF-IDF. Model klasifikasi dibangun menggunakan algoritma Support Vector Machine (SVM), yang dalam penelitian ini dibagi ke dalam tiga skenario pembagian data latih dan uji, yaitu 90:10, 80:20, dan 70:30. Untuk mengatasi kelas yang tidak seimbang, digunakan teknik oversampling ADASYN, SMOTE, dan SMOTETomek. Setiap skenario dikombinasikan dengan tiga kernel SVM, yaitu Linear, RBF, dan Polynomial, sehingga diperoleh total 27 model. Hasil terbaik diperoleh dari kombinasi kernel RBF, metode SMOTE, dan pembagian data 90:10 dengan akurasi mencapai 85%. Penelitian ini juga menghasilkan dashboard berbasis web yang menyajikan hasil analisis secara interaktif dan informatif.

Kata kunci— *Google Play Store, Identitas Kependudukan Digital, Kernel, Resampling, Support Vector Machine*

I. PENDAHULUAN

Transformasi digital dalam sektor pelayanan publik telah memungkinkan masyarakat untuk mengakses layanan secara daring, tidak lagi terbatas pada prosedur manual [1]. Kemajuan pesat dalam bidang teknologi informasi, khususnya dalam sistem identitas digital, telah mendorong instansi pemerintah untuk menerapkan solusi digital supaya meningkatkan kualitas pelayanan. Salah satu bentuk inovasi dalam hal ini adalah peluncuran Identitas Kependudukan Digital (IKD) oleh Direktorat Jenderal Kependudukan dan Pencatatan Sipil (Dukcapil) Kementerian Dalam Negeri. IKD, yang juga dikenal sebagai KTP Digital, bertujuan untuk mendigitalkan dokumen kependudukan, mempercepat integrasi data antar instansi, serta meningkatkan keamanan

identitas melalui sistem verifikasi yang kuat (Peraturan Kemendagri Nomor 72 Tahun 2022).

Kendati demikian, implementasi aplikasi IKD di tengah masyarakat belum sepenuhnya berjalan optimal. Berdasarkan data dari Google Play Store, rating dari aplikasi ini sebesar 3,0 dari lebih dari 65 ribu ulasan, meskipun telah diunduh oleh lebih dari 10 juta pengguna. Banyak pengguna menyampaikan ulasan negatif, terutama terkait kendala teknis, kesulitan dalam proses registrasi, serta kewajiban untuk hadir langsung ke kantor Disdukcapil, yang dianggap tidak sejalan dengan konsep digitalisasi layanan publik. Fenomena ini mengindikasikan adanya kesenjangan antara ekspektasi pengguna dengan kinerja aktual aplikasi.

Dalam konteks tersebut, analisis sentimen menjadi metode yang relevan untuk mengevaluasi opini pengguna terhadap aplikasi IKD. Analisis sentimen atau opinion mining merupakan pendekatan komputasional yang digunakan untuk mengidentifikasi dan mengklasifikasikan emosi atau opini dalam bentuk teks, baik dalam kategori positif, negatif, maupun netral [2]. Hasil analisis ini diharapkan dapat memberikan umpan balik yang konstruktif bagi pengembang aplikasi guna meningkatkan kualitas dan kepuasan pengguna. Berbagai algoritma telah digunakan dalam klasifikasi sentimen, salah satunya adalah Support Vector Machine (SVM), yang dikenal memiliki tingkat akurasi tinggi dan dasar teori yang kuat. Penelitian oleh Rita Ajeng Lestari et al. (2023) menunjukkan bahwa SVM mampu menghasilkan akurasi sebesar 77,0% dalam mengklasifikasikan sentimen pengguna terhadap aplikasi IKD. Selain itu, Azhari et al. (2021) membandingkan performa beberapa algoritma dan menemukan bahwa SVM memberikan akurasi tertinggi, yakni sebesar 95%, dibandingkan dengan algoritma lain seperti C4.5, Random Forest, dan Naïve Bayes.

Berdasarkan latar belakang yang sudah tertera, penelitian ini dilakukan untuk menjawab permasalahan rendahnya tingkat kepuasan pengguna terhadap aplikasi IKD dengan melakukan analisis sentimen berbasis ulasan

pengguna di Google Play Store. Penelitian ini mempunyai tujuan untuk membangun model klasifikasi sentimen menggunakan algoritma SVM, serta mengevaluasi performa model berdasarkan kombinasi skenario pembagian data, metode resampling, dan jenis kernel yang digunakan. Diharapkan hasil penelitian ini dapat berkontribusi nyata terhadap pengembangan aplikasi pelayanan publik berbasis digital yang lebih efektif dan adaptif terhadap kebutuhan pengguna.

II. KAJIAN TEORI

Dasar teori yang digunakan dalam penelitian ini, sebagai landasan konseptual untuk memahami berbagai komponen yang terlibat. Dasar teori menjadi pedoman dalam merancang metode dan pendekatan yang relevan, serta menjadi acuan dalam menginterpretasikan hasil penelitian. Adapun beberapa teori yang mendasari penelitian ini dijelaskan sebagai berikut:

A. Analisis Sentimen

Analisis sentimen merupakan pemrosesan data atau informasi menggunakan komputer yang secara otomatis mengidentifikasi informasi penting dari opini yang disampaikan melalui teks [3]. Bidang ini telah berkembang pesat berkat dampak dan manfaatnya yang signifikan. Kini, banyak perusahaan memanfaatkan analisis sentimen sebagai alat utama untuk menggali wawasan dari beragam data opini [4]. Kita bisa memahami apakah suatu kalimat bernuansa negatif atau positif dengan analisis sentimen [5].

B. Identitas Kependudukan Digital

Direktorat Jenderal Kependudukan dan Pencatatan Sipil Kementerian Dalam Negeri berinovasi menciptakan suatu aplikasi yang dinamakan Identitas Kependudukan Digital (IKD). Hal ini telah diatur dalam Peraturan Menteri Dalam Negeri (Permendagri) Nomor 72 Tahun 2022. IKD merupakan versi digital dari Kartu Tanda Penduduk elektronik, yang juga dikenal sebagai KTP digital. Untuk memperoleh IKD, masyarakat cukup mengunduh aplikasi Identitas Kependudukan Digital melalui perangkat smartphone [6].

C. Scraping

Proses Pengambil dokumen dari internet, seperti halaman Web HTML atau XHTML. Dokumen yang diambil kemudian dianalisis untuk mendapatkan data spesifik yang bisa digunakan untuk berbagai keperluan yang biasa disebut web scraping [7]. Web scraping memiliki tujuan utama, yaitu menemukan dan mengambil informasi dari bagian tertentu suatu web. Ini berbeda dengan web crawling yang menjelajahi seluruh situs beserta semua tautan yang berhubungan. Web scraping hanya fokus pada ekstraksi data tertentu yang dibutuhkan dari situs target

D. Preprocessing

Tahapan awal dalam pengolahan teks adalah Preprocessing. Tujuannya adalah untuk membersihkan dataset agar siap untuk dianalisis [8]. Tahapan ini dilakukan untuk menyaring dan menyederhanakan data sehingga menghasilkan hasil yang relevan dan lebih mudah dikelola oleh sistem [9]. Dalam penelitian ini, tahapan preprocessing

meliputi cleansing, case folding, tokenizing, filtering, normalization, dan stemming.

E. Pembobotan TF-IDF

Pembobotan Term Frequency - Inverse Document Frequency (TF - IDF) adalah proses ekstraksi fitur dengan tujuan memberikan nilai atau bobot pada setiap kata dalam data pelatihan. Untuk mengukur tingkat pentingnya sebuah kata dalam merepresentasikan suatu kalimat, dilakukan proses pembobotan atau perhitungan nilai. Salah satu yang digunakan adalah pembobotan TF-IDF, di mana pemberian nilai didasarkan pada seberapa sering kata itu muncul dalam suatu dokumen dibandingkan dengan seluruh koleksi dokumen [10].

F. SMOTE

Synthetic Minority Over-sampling Technique merupakan metode oversampling yang bertujuan untuk menyeimbangkan distribusi data dengan meningkatkan jumlah sampel pada kelas minoritas. Teknik ini menghasilkan data sintetis baru berdasarkan sampel yang sudah ada hingga jumlahnya setara dengan kelas mayoritas. Meskipun oversampling secara umum berisiko menimbulkan overfitting, SMOTE dirancang untuk meminimalkan risiko tersebut dengan memanfaatkan algoritma K-Nearest Neighbors (KNN) dalam proses pembangkitan data sintetis sesuai dengan jumlah oversampling yang diinginkan [11].

G. ADASYN

Adaptive Synthetic Sampling adalah metode oversampling yang dirancang untuk menyelesaikan permasalahan kelas yang tidak seimbang dalam pembelajaran mesin. Konsep utama dari ADASYN adalah memanfaatkan distribusi bobot berdasarkan tingkat kesulitan pembelajaran pada kelas minoritas. Dengan demikian, ADASYN menghasilkan data sintetis yang lebih banyak dari sampel minoritas yang rumit dipelajari, dibandingkan dengan sampel yang lebih mudah. ADASYN meningkatkan pembelajaran dengan dua acara utama, yaitu pengurangan bias oleh kelas yang tidak seimbang dan secara adaptif menggeser batas keputusan klasifikasi ke arah data yang lebih sulit [12].

H. SMOTE-Tomek Link

Kombinasi SMOTE dan Tomek Links yang diperkenalkan oleh Batista (2003) bertujuan untuk menyeimbangkan data dan mengurangi noise. SMOTE menambahkan data sintetis pada kelas minoritas dengan membuat sampel baru berdasarkan tetangga terdekat. Setelah oversampling, metode Tomek Links digunakan untuk mengidentifikasi pasangan data dari kelas berbeda yang saling berdekatan, lalu menghapus data dari kelas mayoritas untuk memperjelas batas antar kelas. Proses ini diulang hingga dataset bersih dari pasangan yang ambigu [13].

I. Support Vector Machine

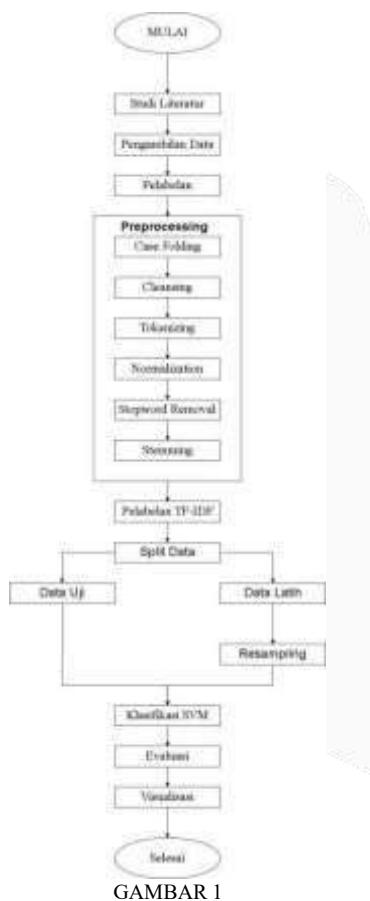
Support Vector Machine adalah algoritma klasifikasi pembelajaran terawasi yang fungsinya memprediksi kelas sebuah data. Prediksi ini dilakukan berdasarkan pola yang sudah dipelajari dari data-data sebelumnya [14]. Prinsip kerja utama dari SVM adalah mencari garis pemisah yang bertujuan untuk memisahkan dua kelas secara optimal.

J. Confusion Matrix

Evaluasi diperlukan untuk mengetahui sejauh mana suatu model mampu mengklasifikasikan data dengan benar. Metode yang umum digunakan dalam evaluasi model adalah confusion matrix. Confusion matrix merupakan sebuah matriks yang menyajikan informasi mengenai jumlah prediksi yang benar dan salah pada data uji, sehingga dapat digunakan untuk menilai kualitas kinerja model klasifikasi tersebut [15]. Melalui confusion matrix, kita dapat mengevaluasi performa machine learning berdasarkan berbagai metrik seperti akurasi, presisi, recall, spesifikasi, dan F1 Score.

III. METODE

Rangkaian penyelesaian masalah dalam penelitian ini disusun secara struktur guna mencapai tujuan yang telah ditetapkan. Proses ini dimulai dari studi literatur hingga tahap visualisasi. Setiap tahapan saling berkaitan dan berkontribusi dalam bentuk model analisis sentimen yang optimal. Diagram berikut menyajikan alur keseluruhan proses penelitian ini:



GAMBAR 1
(ALUR PENYELESAIAN MASALAH)

A. Studi Literatur

Studi Literatur merupakan serangkaian kegiatan untuk mencari penelitian penelitian terdahulu yang serupa dimana informasi yang ada pada penelitian terdahulu akan dijadikan landasan ataupun acuan dalam pelajaran penelitian. Studi literatur sendiri dapat ditemukan dengan mempelajari dan menganalisis dari beberapa sumber referensi yaitu: studi kasus, artikel ilmiah, paper, buku, maupun jurnal lama ataupun media jurnal terbaru yang tentunya tersedia secara

gratis maupun berbayar. Jurnal dan artikel dalam penelitian ini diambil dari platform Google Scholar.

B. Pengumpulan Data

Peneliti mengambil data dari ulasan – ulasan yang berada pada aplikasi IKD pada Play Store menggunakan teknik scraping. Scraping adalah teknik pengambilan data di sebuah platform. Hasil dari scraping didapat sebanyak 10.000 data ulasan berbahasa indonesia dengan rentang waktu dari 1 Agustus 2023 – 31 Januari 2025. Data yang diambil akan diubah menjadi CSV untuk mempermudah proses ke tahap berikutnya. Hasil pengambilan data dapat dilihat pada gambar 3.2.

C. Pelabelan

Sebelum masuk ke tahap preprocessing dilakukan pelabelan secara manual dengan tujuan untuk menghasilkan dataset yang akurat dan handal. Pelabelan dilakukan oleh sejumlah relawan yang telah diberikan arahan. Setiap teks diberi label berdasarkan kategori sentimen, yaitu positif dan negatif. Kategori “positif” diberikan jika teks menunjukkan kepuasan atau dukungan, “negatif” jika menunjukkan ketidakpuasan atau kritikan

D. Preprocessing

Preprocessing dilakukan untuk membersihkan dan menyeragamkan data teks guna meningkatkan akurasi dan efektivitas dalam analisis sentimen. Tahapan ini bertujuan untuk mengurangi elemen-elemen yang tidak relevan sehingga model dapat mempelajari pola secara lebih optimal.

Terdapat beberapa tahapan pada preprocessing, yaitu :

Cleansing

: Tahap ini berujuan untuk menghilangkan gangguan dalam teks seperti symbol-simbol yang tidak memiliki makna dalam kontek analisis sentimen.

Case Folding

: Tahap ini bertujuan untuk menyeragamkan format teks dengan cara merubah seluruh teks menjadi huruf kecil (lowercase)

Tokenizing

: Tokenizing merupakan proses pemecahan teks atau kalimat menjadi unit-unit yang lebih kecil, yang disebut token.

Normalization

: Normalization adalah proses untuk mengubah kata tidak baku ditingkatkan menjadi kata formal.

Stopword Removal

: Stopword Removal dilakukan untuk menghilangkan kata tidak penting dan tidak memiliki arti

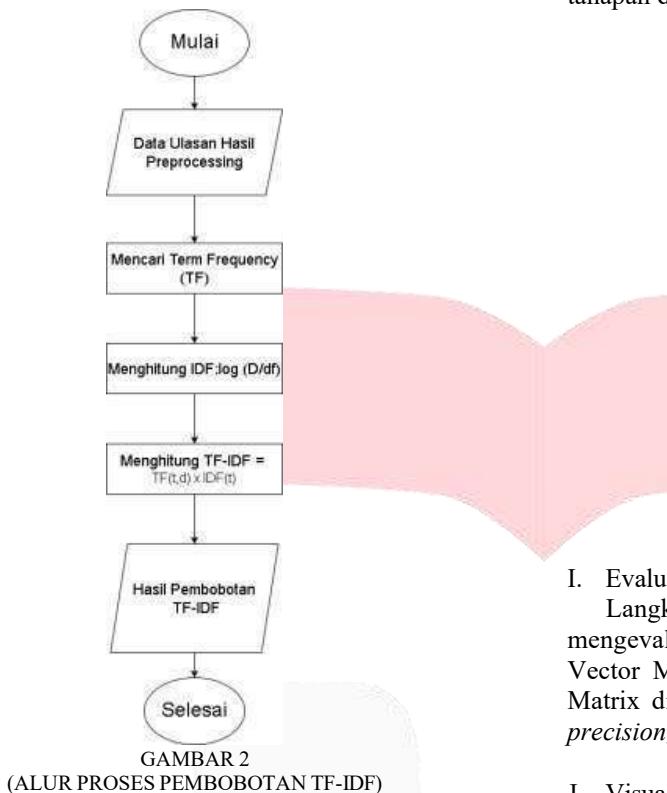
Stemming

: Stemming dilakukan untuk mengubah semua menjadi kata dasar

E. Pembobotan TF-IDF

Setelah melalui proses pelabelan maka akan dilakukan pembobotan kata menggunakan TF-IDF (Term Frequency - Inverse Document Frequency) Untuk mengetahui seberapa penting sebuah kata mewakili sebuah kalimat. Dalam prosesnya, dilakukan perhitungan jumlah kemunculan tiap

term pada keseluruhan data ulasan (TF), menghitung jumlah ulasan yang mengandung setiap term (DF), menghitung inverse dari nilai DF (IDF), dan mengalikan nilai tf dan idf untuk memperoleh nilai bobot tiap ulasan. Alur proses pembobotan bisa dilihat pada gambar (2).



F. Split Data

Pada tahap ini, pembagian data dilakukan untuk menentukan data latih dan data uji untuk memastikan model memiliki set data terpisah untuk pelatihan, validasi dan pengujian. Pada penelitian ini pembagian dilakukan dengan `split_test 0.1, 0.2, dan 0.3` yang berarti data dibagi menjadi tiga skenario, yaitu 90% data latih 10% data uji, 80% data latih 20% data uji, 70% data latih 30% data uji

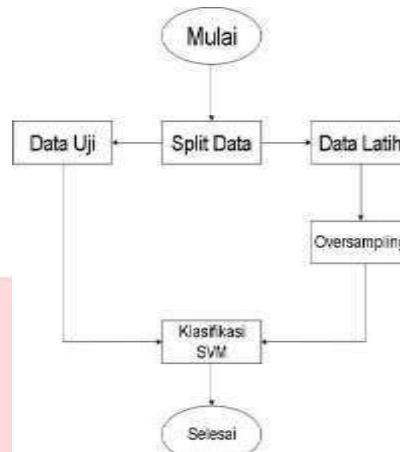
G. Resampling

Tahap ini dilakukan untuk penanganan terhadap ketidakseimbangan data dengan tujuan menyeimbangkan kelas positif dan negatif. Metode resampling yang digunakan dalam penelitian ini, yaitu SMOTE, ADASYN, SMOTETomek. Dalam penelitian ini, penerapan metode resampling hanya digunakan pada data latih.

H. Klasifikasi SVM

Setelah dilakukan penyeimbangan data menggunakan metode resampling, langkah selanjutnya yaitu menerapkan fungsi kernel guna memetakan data ke ruang fitur berdimensi lebih tinggi. Fungsi ini berperan penting dalam mempermudah pemisahan kelas oleh algoritma SVM. Jenis kernel yang dapat digunakan antara lain Linear, Polynomial, serta *Radial Basis Function* (RBF). Kemudian, proses pelatihan model dilakukan dengan memanfaatkan data latih. Pada tahap ini, algoritma Support Vector Machine akan

menentukan hyperplane optimal yang mampu memisahkan kelas secara maksimal. Setelah pelatihan selesai, model diuji menggunakan data uji yang sebelumnya tidak digunakan dalam proses pelatihan. Alur proses klasifikasi menggunakan SVM ini dapat dilihat pada Gambar (3), yang menjelaskan tahapan dari pembagian data.



I. Evaluasi

Langkah selanjutnya setelah klasifikasi adalah mengevaluasi untuk menilai sejauh model dari Support Vector Machine (SVM) dapat mengklasifikasi. Confusion Matrix dilakukan untuk mengevaluasi nilai dari *accuracy*, *precision*, *recall*, dan *f1-score*.

J. Visualisasi

Pada penelitian ini melakukan visualisasi data berupa dashboard dari web yang dibuat menggunakan library streamlit dengan Bahasa pemrograman python. tersebut memiliki dua menu. Dibawah ini merupakan mockup dari perancangan dashboard yang akan dibuat. Menu yang pertama berisikan diagram diagram yang dibuat dengan tools *Looker Studio*. Menu yang kedua berisikan visualisasi data yaitu word cloud yang dibuat dengan tools tools yang tersedia.

IV. HASIL DAN PEMBAHASAN

BAB ini menyajikan paparan objektif terhadap hasil yang diperoleh dari proses penelitian. Penjelasan difokuskan pada analisis terhadap data yang telah dikumpulkan, termasuk temuan-temuan penting yang berhasil diidentifikasi. Proses interpretasi data dilakukan secara sistematis guna memperoleh hubungan yang relevan, serta menghasilkan kesimpulan atau generalisasi dari temuan yang ada.

A. Pengambilan Data

Pengumpulan data dilakukan melalui metode web scraping terhadap ulasan pengguna aplikasi IKD yang tersedia di Google Play Store. Proses pengambilan data di ambil dalam jangka waktu 1 Agustus 2023 hingga 31 Januari 2025, dan berhasil memperoleh 15.855 ulasan dengan kolom kolom 'userName', 'score', 'at', 'content'.

Setelah data terkumpul, langkah berikutnya adalah melakukan seleksi untuk menyaring ulasan-ulasan yang tidak

relevan. Ulasan yang bersifat tidak bermakna, bersifat spam, dan tidak terkait dengan pengalaman pengguna aplikasi IKD dihapus dari dataset. Dari proses seleksi ini diperoleh 1.191 ulasan yang dianggap valid dan layak untuk diproses lebih lanjut.

Tahap seleksi ini penting untuk menjamin bahwa data yang dianalisis benar-benar mencerminkan persepsi dan pengalaman dari pengguna aplikasi IKD, sehingga dapat meningkatkan keakuratan dan relevansi hasil. Rincian hasil scraping dapat dilihat pada tabel berikut.

TABEL 1
(HASIL SCRAPING)

no	userName	score	at	content
0	Pengguna Google	5	1/31/2025 8:47	Masa akun ke blokir setelah pembaruan/upgrade aplikasi
...
1191	Pengguna Google	1	8/1/2023 4:00	Tolong diperbaiki, karena tidak ada pilihan cetak atau download dokumen.

B. Labeling

Dalam tahap labeling ini menggunakan labeling secara manual. Dalam pelaksanaanya penelitian ini menggunakan satu orang relawan untuk melakukan pelabelan secara manual. Dalam pelaksanaanya data akan diberikan ke relawan lalu mereka akan membaca setiap kalimat dalam data ulasan dan memberi label positif atau negatif pada data ulasan menurut perspektif mereka. Cara ini bertujuan untuk memastikan keakuratan tinggi pada data yang membutuhkan pemahaman kontekstual. Berikut adalah hasil dari labeling secara manual.

TABEL 2
(LABELING MANUAL)

no	userName	content	value
0	Pengguna Google	Aplikasi ini jga luarbiasa, ketika saya butuh identitas, ternyata ktp tdk ada di dompet, tinggal buka aplikasi aja..udah deh	Positif
...
1191	Pengguna Google	Tolong diperbaiki, karena tidak ada pilihan cetak atau download dokumen.	Negatif

C. Preprocessing

Proses ini merupakan tahapan awal dari pengolahan data melibatkan persiapan dan pembersihan data mentah untuk memastikan kualitas data yang optimal. Hal ini penting untuk menghilangkan data yang tidak relevan dalam dataset. Pada proses ini memiliki beberapa tahapan sebelum menuju ke tahap klasifikasi, yaitu case folding, cleansing text, tokenizing, normalization, stopword removal, stemming. Setelah melakukan proses preprocessing, memiliki hasil yang berbeda dengan yang sebelum dilakukannya preprocessing. Hasil yang didapat saat sebelum preprocessing yaitu berkisar

1191 dataset dan hasil yang didapat setelah melakukan proses preprocessing yaitu 1191 dataset. Tahapan preprocessing bisa dilihat pada tabel dibawah ini.

Tahap pertama ada tahapan case folding dalam tahap ini, akan menyamakan semua huruf dalam dataset menjadi huruf kecil (lowercase). Hal ini dilakukan agar proses analisis seperti pencocokan kata atau pembagian kata menjadi token, dapat dilakukan secara efisien. Dengan proses ini, kata seperti "Data", "DATA", "data" akan dianggap sebagai kata yang sama, sehingga meminimalisir redundansi dan meningkatkan akurasi dalam pengolahan data.

TABEL 3
(CASE FOLDING)

Sebelum	Sesudah
Aplikasi ini jga luarbiasa, ketika saya butuh identitas, ternyata ktp tdk ada di dompet, tinggal buka aplikasi aja..udah deh	aplikasi ini jga luarbiasa, ketika saya butuh identitas, ternyata ktp tdk ada di dompet, tinggal buka aplikasi aja..udah deh

Tahapan kedua ada tahap cleansing text pada tahap ini, membersihkan text yang tidak relevan, berlebihan, dan tidak diperlukan dalam analisa untuk menghasilkan teks yang lebih bersih dan siap dianalisis lebih lanjut. Beberapa langkah yang ada dalam tahap ini, seperti menghapus emoji, tanda baca, angka, spasi yang berlebih. Karakter karakter khusus tersebut dapat mempengaruhi pemrosesan teks dan menambah kompleksitas dalam ekstraksi informasi.

TABEL 4
(CLEANSING TEXT)

Sebelum	Sesudah
aplikasi nya ngebug setelah swafoto tidak bisa melanjutkan ke tahap selanjutnya m	aplikasi nya ngebug setelah swafoto tidak bisa melanjutkan ke tahap selanjutnya

Tahapan ketiga ada tahap tokenizing, tahap untuk memecah teks menjadi unit-unit terkecil (token) untuk analisis lebih lanjut. Hal ini dilakukan untuk membuat teks menjadi terstruktur dan memberikan input yang sesuai untuk algoritma. Tokenizing ini menjadi kunci untuk membuka potensi analisis text, dari hitung kata hingga pemodelan bahasa yang kompleks. Berikut merupakan hasil dari tahapan tokenizing menggunakan python string split () pada Tabel 5.

TABEL 5
(TOKENIZING)

Sebelum	Sesudah
bagaimana caranya minta barcode tanpa harus ke dukcapil atau ke fasyankes tingkat	[bagaimana, caranya, minta, barcode, tanpa, harus, ke, dukcapil, atau, ke, fasyankes, tingkat]

Tahapan keempat ada tahap normalization, tahap ini memproses teks mentah menjadi bentuk yang lebih terstruktur dan konsisten, sehingga memudahkan untuk mengidentifikasi pola dan informasi penting di dalamnya. Dengan normalization, kita bisa memastikan kata sebelumnya seperti "saya", "sy", "Sya" menjadi "saya".

TABEL 6
(NORMALIZATION)

Sebelum	Sesudah
[masuk,harus,login,lg,setelah,update,jelek,aplikasi,yg,sekarang,gak,kaya,dulu,jelek,pol,jelek]	masuk harus login lagi setelah update jelek aplikasi yang sekarang enggak kayak dulu jelek pol jelek

Tahapan kelima ada tahap stopword removal, tahap ini bertujuan untuk menghapus kata-kata yang tidak memiliki makna penting atau tidak memberikan kontribusi signifikan terhadap analisa. Kata-kata yang termasuk stopword biasanya merupakan kata imbuhan seperti ‘dan’, ‘di’, ‘yang’, yang akan dibandingkan dengan daftar stopword. Kata-kata yang teridentifikasi sebagai stopword akan dihapus dari teks.

TABEL 7
(STOPWORD REMOVAL)

Sebelum	Sesudah
aplikasi ini juga luarbiasa ketika saya butuh identitas ternyata ktp tidak ada di dompet tinggal buka aplikasi ajaudah deh	aplikasi luarbiasa butuh identitas ktp dompet tinggal buka aplikasi ajaudah deh

Tahapan keenam ada tahap Stemming, merupakan tahapan mengubah kata menjadi bentuk dasar, dengan menghilangkan imbuhan atau akhiran tertentu seperti, awalan, akhiran, atau sisipan. Tujuan dari tahap ini mencocokkan kata-kata yang sebenarnya sama artinya, meskipun tulisannya berbeda. Berikut adalah hasil proses stemming dengan menggunakan library sastrawi.

TABEL 8
(STEMMING)

Sebelum	Sesudah
pembaharuan aplikasi ikd pas coba membuka nya akses perizinan menyeting semuannya setuju masuk aplikasi nya mohon solusi nya ikd kebuka aplikasi ikd beguna perantauan akses pakai ktp digital	baru aplikasi ikd pas coba buka nya akses izin ting semu tuju masuk aplikasi nya mohon solusi nya ikd buka aplikasi ikd beguna rantau akses pakai ktp digital

D. Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan sebuah metode pembobotan yang sering dijumpai untuk mengevaluasi seberapa penting suatu kata terhadap dokumen tertentu dalam sekumpulan dokumen (korpus). Hasil dari tahapan preprocessing akan dihitung menggunakan pembobotan TF-IDF untuk mendapatkan nilai frekuensi relatif dari setiap kata. Nilai-nilai ini menggambarkan kontribusi masing-masing kata terhadap isi dokumen dan dapat digunakan sebagai fitur dalam proses klasifikasi atau analisis lebih lanjut. Berikut ini ditampilkan hasil frekuensi kata berdasarkan pembobotan TF-IDF dari data ulasan yang telah diproses.

TABEL 9
(FREKUENSI KATA)

no	term	DF
0	"aplikasi"	590
1	"update"	242
2	"daftar"	206
...
2891	"surabaya"	1
2892	"samsung"	1
2893	"salamsehat"	1

Tabel 9 menyajikan hasil perhitungan Document Frequency (DF) terhadap setiap term dalam data ulasan. Dari hasil tersebut terlihat bahwa kata “aplikasi” memiliki frekuensi paling tinggi, yaitu 590. Kata-kata lain seperti “data”, “update”, “daftar”, “mudah” juga termasuk yang sering muncul. Ini mengidentifikasi bahwa kata-kata tersebut muncul secara berulang dalam ulasan. Kata-kata ini biasanya memiliki informasi yang rendah dalam proses klasifikasi, karena kurang mampu membedakan antar kategori.

Sebaliknya, terdapat kata-kata seperti “gegara”, “surabaya”, “samsung”, “salamsehat” yang hanya muncul satu kali di seluruh ulasan. Meskipun jarang muncul, kata-kata ini justru bisa memiliki nilai informasi yang tinggi. Term dengan frekuensi rendah membantu dalam membedakan antar kelas pada saat proses klasifikasi.

E. Klasifikasi Support Vector Machine

Tahapan ini, membangun model menggunakan algoritma Support Vector Machine (SVM) dengan melakukan pembagian data terlebih dahulu menjadi data latih dan data uji. Pemodelan dilakukan dengan menerapkan tiga skenario `test_size`, yaitu 0.1, 0.2, dan 0.3. Nilai-nilai ini mempresentasikan proporsi data uji terhadap keseluruhan dataset, sehingga secara otomatis menentukan proporsi data latih masing-masing sebesar 90%, 80%, dan 70%. Tujuannya adalah untuk mengevaluasi performa model klasifikasi dari macam-macam skenario jumlah data latih dan data uji dan digunakan untuk mengevaluasi sejauh mana model mampu melakukan generalisasi terhadap data yang belum pernah dilihat sebelumnya. Evaluasi ini penting guna memastikan bahwa model tidak hanya bekerja baik pada data latih, tetapi juga memiliki kinerja yang andal pada data uji.

Setelah melakukan tahapan pembagian data, langkah selanjutnya melakukan teknik oversampling guna menangani ketidakseimbangan data, di mana terdapat perbedaan jumlah data yang signifikan antara satu kelas dengan kelas lainnya. Dalam penelitian ini, digunakan teknik oversampling seperti SMOTE, ADASYN, dan SMOTETomek guna menangani ketidakseimbangan data. Penerapan teknik ini dilakukan pada data latih dari tiap `test_size`, untuk meningkatkan proporsi data kelas minoritas dalam proses pelatihan model, tanpa mempengaruhi data uji yang digunakan untuk evaluasi performa. Informasi mengenai distribusi data latih sebelum oversampling tercantum pada Tabel 10.

TABEL 10
(DISTRIBUSI DATA LATIH SEBELUM OVERSAMPLING)

Kelas	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Positif	174	174	174
Negatif	897	897	897

Dengan menggunakan metode SMOTE, ADASYN, dan SMOTETomek, data dapat diseimbangkan agar tidak terjadi bias pada model. Setelah dilakukan penyeimbangan pada data latih dengan metode SMOTE, didapatkan hasil distribusi dengan hasil sempurna antar kelas positif dan negatif di semua rasio. Pada rasio 90:10 berjumlah 897 data dari tiap kelas, rasio 80:20 berjumlah 798 data dari tiap kelas, dan rasio 70:30 berjumlah 698 data dari tiap kelas. Hal ini menunjukkan bahwa SMOTE telah melakukan oversampling terhadap kelas minoritas hingga setara dengan kelas mayoritas tanpa menghapus data dari kelas manapun. Distribusi data latih dengan metode SMOTE disajikan pada tabel Tabel 11.

TABEL 11
(DISTRIBUSI DATA LATIH SETELAH SMOTE)

Kelas	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Positif	897	798	698
Negatif	897	798	698

Dalam penerapan metode ADASYN menunjukkan bahwa jumlah data pada kelas positif tetap lebih sedikit dibandingkan dengan kelas negatif. Pada rasio 90:10, jumlah data positif adalah 850, sedangkan data negatif sebanyak 897. Hal serupa juga terjadi pada rasio 80:20 (747 positif dan 798 negatif) dan pada rasio 70:30 (671 positif dan 698 negatif). Ini menunjukkan bahwa ADASYN belum sepenuhnya menyamakan distribusi antar kelas, namun kemungkinan hanya menyesuaikan hingga tingkat ketidakseimbangan minimum. Distribusi data latih dengan metode ADASYN disajikan pada Tabel 12.

TABEL 12
(DISTRIBUSI DATA LATIH SETELAH ADASYN)

Kelas	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Positif	850	747	671
Negatif	897	798	698

Distribusi dalam metode SMOTETomek sama dengan metode SMOTE yang mendistribusikan kelas secara sempurna antara kelas positif dan negatif. Pada rasio 90:10 berjumlah 897 data pada masing-masing kelas. Rasio 80:20 dan 70:30, masing-masing kelas memiliki 798 dan 689 data. Distribusi data latih dengan metode SMOTETomek disajikan pada Tabel 13.

TABEL 13
(DISTRIBUSI DATA LATIH SETELAH SMOTETOMEK)

Kelas	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Positif	897	798	698
Negatif	897	798	698

Hal selanjutnya setelah melakukan oversampling adalah klasifikasi menggunakan metode Support Vector Machine (SVM). Dalam penelitian ini, klasifikasi menggunakan tiga kernel yaitu Linear, RBF, dan Polynomial. Tujuannya adalah untuk mengevaluasi dan membandingkan performa masing-masing dalam menghasilkan akurasi klasifikasi yang optimal. Proses Evaluasi dilakukan dengan menguji `test_size` pada

data latih yang telah melalui proses oversampling, guna memperoleh tingkat akurasi yang paling tinggi. Perolehan nilai proses klasifikasi disajikan pada tabel-tabel di bawah.

TABEL 14
(HASIL AKURASI DARI OVERSAMPLING ADASYN)

Kernel	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Linear	0,82	0,84	0,84
RBF	0,85	0,84	0,84
Polynomial	0,42	0,40	0,28

Tabel 14 menunjukkan bahwa kernel RBF memberikan hasil akurasi yang tertinggi, dengan nilai yaitu 0,85 pada rasio 90:10 dan nilai 0,84 pada rasio 80:20 dan 70:30. Pada kernel Linear juga menunjukkan akurasi yang tinggi, dengan akurasi sebesar 0,82 pada rasio 90:10 dan 0,84 pada rasio 80:20 dan 70:30, yang menunjukkan bahwa model ini mampu mengklasifikasikan data hasil oversampling.

ADASYN dengan cukup efektif. Sebaliknya, kernel Polynomial memiliki akurasi yang rendah, dengan akurasi sebesar 0,42 pada rasio 90:10, 0,40 pada rasio 80:20, dan 0,28 pada rasio 70:30, yang menunjukkan bahwa kernel ini kurang cocok digunakan dalam data hasil oversampling ADASYN. Secara keseluruhan, kernel RBF menjadi pilihan dalam mengklasifikasi data hasil oversampling ADASYN.

TABEL 15
(HASIL AKURASI DARI OVERSAMPLING SMOTE)

Kernel	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Linear	0,84	0,83	0,84
RBF	0,85	0,84	0,85
Polynomial	0,42	0,34	0,30

Pada Tabel 15 Menunjukkan bahwa kernel RBF masih memberikan hasil akurasi yang tertinggi, dengan nilai 0,85 pada rasio 90:10 dan 70:30, serta nilai 0,84 pada rasio 80:20. Kernel Linear mendapatkan hasil akurasi yang tertinggi, dengan akurasi sebesar 0,84 pada rasio 90:10 dan 70:30, nilai 0,83 pada rasio 80:20, yang menunjukkan bahwa model ini cukup efektif digunakan pada data hasil oversampling SMOTE. Sementara itu, kernel Linear menunjukkan hasil yang kurang maksimal dengan akurasi hanya 0,42 pada rasio 90:10, 0,34 pada rasio 80:20, dan 0,30 pada rasio 70:30. Secara keseluruhan, kernel RBF masih menjadi pilihan dalam mengklasifikasi data hasil oversampling SMOTE

TABEL 16
(HASIL AKURASI DARI OVERSAMPLING SMOTETOMEK)

Kernel	Skenario Pembagian Data Latih		
	90:10	90:10	90:10
Linear	0,84	0,83	0,84
RBF	0,85	0,84	0,85
Polynomial	0,42	0,34	0,30

Tabel 16 menunjukkan bahwa kernel RBF tetap memberikan hasil akurasi yang tertinggi, dengan nilai 0,85 pada rasio 90:10 dan 70:30, serta 0,84 pada rasio 80:20. Pada kernel Linear juga memberikan hasil yang tinggi, dengan akurasi bernilai 0,84 pada rasio 90:10 dan 70:10, serta 0,83

pada rasio 80:20, hal ini memperlihatkan bahwa kernel ini menjadi pilihan untuk data hasil oversampling

SMOTETomek. Sementara itu, kernel Polynomial menunjukkan hasil yang rendah. Nilai akurasi berkisar 0,42 pada rasio 90:10, 0,34 pada rasio 80:20, dan 0,30 pada rasio 70:30. Secara keseluruhan, kernel RBF masih menjadi pilihan dalam klasifikasi data hasil oversampling SMOTETomek.

Tabel-tabel diatas merupakan hasil klasifikasi dari berbagai jenis kernel SVM pada tiap rasio pembagian data (test_size) pada data latih yang telah di oversampling. Berdasarkan hasil tersebut, dapat disimpulkan bahwa kernel RBF memberikan performa terbaik dengan nilai akurasi tertinggi mencapai 0,85 secara konsisten. Sementara itu, kernel Linear memiliki akurasi yang sedikit lebih rendah yaitu 0,84. Di sisi lain, kernel Polynomial menghasilkan akurasi yang paling rendah, dengan rata rata hanya sekitar 0,42, sehingga kurang direkomendasikan untuk digunakan dalam kasus ini.

F. Evaluasi Model

Tahap ini menyajikan hasil evaluasi dari proses analisis sentimen yang dilakukan menggunakan algoritma SVM. Dalam analisis ini, data dikategorikan ke dalam dua jenis sentimen, yaitu positif dan negatif. Detail hasil klasifikasi dari model SVM dengan kernel RBF disajikan pada gambar dibawah ini.

--- Kernel: rbf ---

```
== Hasil Klasifikasi ==
      precision    recall  f1-score   support
Negatif       0.85      0.99      0.92      101
Positif       0.67      0.11      0.18       19

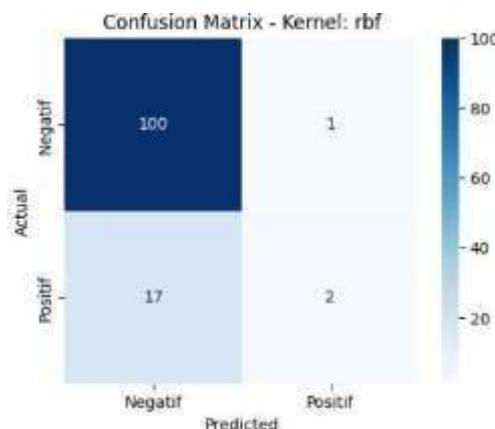
   accuracy                           0.85      120
  macro avg       0.76      0.55      0.55      120
weighted avg    0.82      0.85      0.80      120
```

Akurasi (Kernel: rbf): 0.8500

GAMBAR 4
(CLASSIFICATION REPORT)

Gambar 4, yang menampilkan Classification Report model, menunjukkan model mempunyai kinerja sangat baik dalam klasifikasi kedua sentimen, yaitu positif dan negatif. Pada kelas negatif, model meraih nilai presisi 0.85, recall 0.99, dan f1-score 0.92. Sementara itu, untuk kelas positif, model mencapai presisi 0.67, recall 0.11, dan f1-score 0.18. Secara keseluruhan, model menunjukkan akurasi 0.85, yang menandakan tingkat keberhasilan tinggi dalam mengklasifikasikan seluruh data uji.

Confusion matrix digunakan untuk memahami lebih lanjut distribusi model terhadap masing-masing kelas. Matriks ini memperlihatkan detail yang lebih jelas tentang prediksi yang benar maupun yang salah, sehingga membantu untuk mengidentifikasi ketidakseimbangan dan kelemahan dalam performa model. Gambar 5 memperlihatkan bahwa Confusion matrix mengklasifikasikan data dengan tingkat akurasi yang tinggi pada kelas negatif. Hal ini ditunjukkan oleh jumlah prediksi benar terhadap kelas negatif sebanyak 100 data, dan 1 data negatif yang salah sebagai data positif. Namun, dalam klasifikasi kelas positif menunjukkan jumlah prediksi benar terhadap kelas positif sebanyak 19 data, dan 2 data yang benar positif, sementara 17 data lainnya justru salah diklasifikasikan sebagai negatif.



GAMBAR 5
(CONFUSION MATRIX)

G. Visualisasi

Pada bagian ini, visualisasi berupa web menggunakan framework streamlit. Visualisasi ini berisikan hasil dari proses langkah-langkah saat analisis sentimen. Pada Gambar 6 diperlihatkan tampilan visualisasi yang telah dikembangkan menggunakan framework streamlit.

Identitas Kependudukan Digital

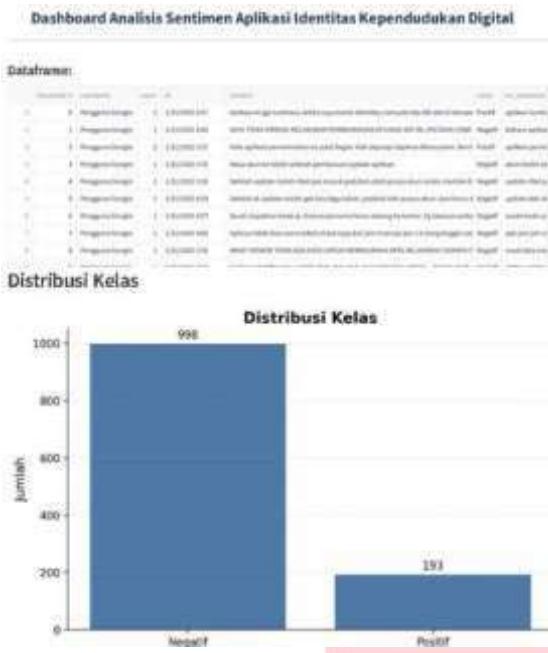


Latar Belakang Aplikasi

Pelayanan publik kini berkembang pesat ke arah digital. Salah satunya melalui Identitas Kependudukan Digital (IKD), layanan dari Birojen Dekapitil Kemendagri yang memungkinkan masyarakat mengakses dokumen Kependudukan secara digital melalui aplikasi mobile. Namun, rating aplikasi IKD di Play Store masih rendah (3.5) dan banyak pengguna menyampaikan ketuhanan. Oleh karena itu, dilakukan analisis sentimen terhadap ulasan pengguna untuk mengetahui persepsi dan fokus perbaikan aplikasi. Dalam penelitian ini digunakan metode Support Vector Machine (SVM) karena dikenal memiliki akurasi tinggi dan performa baik dalam klasifikasi sentimen.

GAMBAR 6
(TAMPILAN HALAMAN UTAMA)

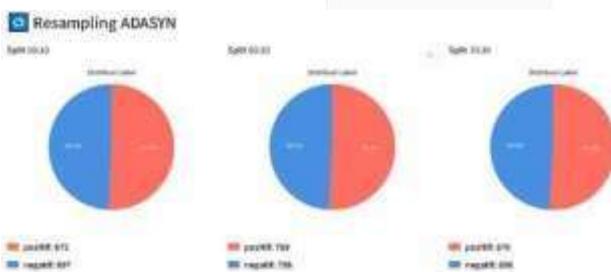
Pada Gambar 6 memperlihatkan tampilan awal web yang memperlihatkan tampilan yang sederhana. Navigasi utama berada di sidebar sebelah kiri yang memudahkan pengguna untuk mengakses halaman lain seperti Home, Dashboard, dan Analisis Sentimen. Pada halaman utama yaitu Home, menampilkan judul aplikasi "Identitas Kependudukan Digital" yang dilengkapi logo aplikasi tersebut untuk memperkuat identitas visualnya. Dibawahnya, terdapat latar belakang aplikasi yang memberikan gambaran umum mengenai tujuan dan kegunaan aplikasi.



GAMBAR 7
(VISUALISASI DISTRIBUSI KELAS PADA DASHBOARD)

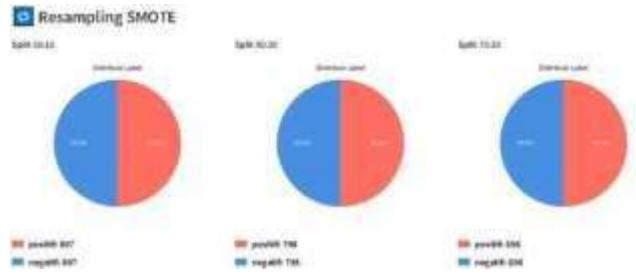
Gambar 7 menampilkan bagian kedua dari web yaitu ‘Dashboard’, menunjukkan tampilan data hasil dari proses preprocessing. Dalam data tersebut menunjukkan kolom “username”, “score”, “at”, “content”, dan “cleaned_text”.

Lalu ada visualisasi menggunakan diagram batang untuk menunjukkan jumlah ulasan dari kelas positif dan negatif. Hasil yang diperoleh menunjukkan mayoritas ulasan ada di kelas positif dengan jumlah 998 ulasan, kelas positif berjumlah 193 ulasan.



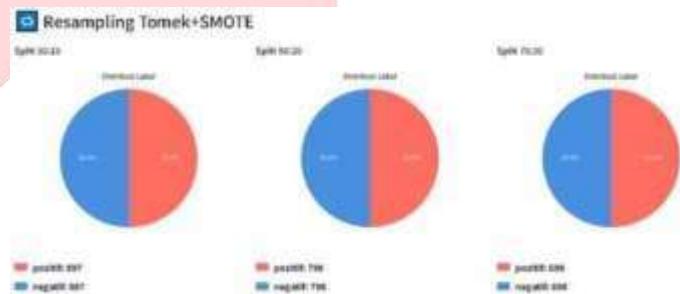
GAMBAR 8
(VISUALISASI HASIL DISTRIBUSI OVERSAMPLING METODE ADASYN)

Gambar 8 merupakan visualisasi distribusi kelas setelah dilakukan proses oversampling menggunakan metode ADASYN pada tiga skenario pembagian data, yaitu 90:10, 80:20, dan 70:30. Berdasarkan hasil visualisasi, dapat dilihat bahwa pada setiap skenario pembagian data, distribusi label antara kelas positif dan negatif hampir seimbang. Pada split 90:10, data positif berjumlah 872 dan negatif 897; pada split 80:20, data positif sebanyak 769 dan negatif 798; sedangkan pada split 70:30, data positif berjumlah 670 dan negatif 698.



GAMBAR 9
(VISUALISASI HASIL DISTRIBUSI OVERSAMPLING METODE SMOTE)

Gambar 9 menunjukkan visualisasi distribusi label setelah diterapkannya metode oversampling SMOTE pada tiga skenario pembagian data, yaitu split 90:10, 80:20, dan 70:30. Berdasarkan hasil visualisasi, terlihat bahwa SMOTE berhasil menghasilkan distribusi label yang sama persis antara kelas positif dan negatif pada setiap skenario split. Pada split 90:10, jumlah data positif dan negatif masing-masing sebanyak 897; pada split 80:20, masing-masing 798; dan pada split 70:30, masing-masing 698.



GAMBAR 10
(VISUALISASI HASIL DISTRIBUSI OVERSAMPLING SMOTETOMEK)

Gambar 10 memperlihatkan hasil visualisasi distribusi label setelah diterapkannya kombinasi metode SMOTETomek pada skenario pembagian data, yaitu split 90:10, 80:20, dan 70:30. Berdasarkan visualisasi, pembagian data kelas negatif dan positif terlihat sama persis pada semua skenario pembagian data. Pada split 90:10, jumlah data positif dan negatif masing-masing sebanyak 897, split 80:20 masing-masing sebanyak 798, dan split 70:30 masing-masing sebanyak 698.



GAMBAR 11
(TABEL CLASSIFICATION REPORT DAN VISUALISASI WORDCLOUD)

Gambar 11 menunjukkan bagian lanjutan dari halaman dashboard, dimana evaluasi perfoma model dilakukan menggunakan metrik klasifikasi standar yaitu precision,

recall, f1-score, dan akurasi. Model evaluasi ini mengambil dari kernel RBF dengan resampling dari SMOTE menggunakan rasio 90:10. Hasil evaluasi menunjukkan precision bernilai 82%, recall bernilai 85%, f1-score bernilai 80%, dan akurasi bernilai 85% yang menandakan performa cukup baik dalam klasifikasi. Lalu, eksplorasi data teks dilakukan melalui visualisasi wordcloud dan daftar 10 kata yang sering muncul. Kata-kata seperti “aplikasi”, “update”, “ktp”, “daftar”, “data”, “dukcapil”, “mudah”, “masuk”, “foto”, “online” muncul paling sering, dengan kata “aplikasi” memiliki frekuensi tertinggi yaitu 590 kali.



GAMBAR 12

(TAMPILAN HALAMAN WEB ANALISIS SENTIMEN)

Gambar 12 merupakan halaman dari analisis sentimen, menyajikan fitur pengujian kalimat yang telah disediakan. Pengguna dapat menuliskan satu kalimat opini mengenai aplikasi ini, lalu dengan menekan uji kalimat maka akan diketahui hasil klasifikasi sentimennya. Fitur ini memudahkan dalam mengevaluasi apakah opini tersebut tergolong positif atau negatif dengan sistem klasifikasi yang telah dibangun sebelumnya. Pengguna dapat melakukan pengujian secara real-time, sehingga menambah nilai praktis dari sistem yang dikembangkan.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa model analisis sentimen terhadap ulasan pengguna aplikasi Identitas Kependudukan Digital (IKD) berhasil dikembangkan dengan menggunakan algoritma Support Vector Machine (SVM). Proses pembangunan model dimulai dari tahap preprocessing data teks, yang mencakup cleansing, case folding, tokenizing, normalization, stopword removal, dan stemming. Selanjutnya, dilakukan pembagian data latih dan data uji serta penerapan teknik oversampling seperti SMOTE, ADASYN, dan SMOTETomek untuk mengatasi ketidakseimbangan kelas. Evaluasi performa klasifikasi menunjukkan bahwa metode SVM mampu memberikan hasil yang cukup baik. Dari 27 model yang diuji, model terbaik diperoleh pada skenario pembagian data 90:10 dengan kernel RBF dan metode oversampling SMOTE, yang menghasilkan nilai precision sebesar 0,82, recall 0,85, f1-score 0,80, dan akurasi 0,85. Hal ini menunjukkan bahwa metode SVM efektif dalam mengklasifikasikan sentimen pengguna terhadap aplikasi IKD. Selain itu, penelitian ini juga menghasilkan visualisasi dalam bentuk dashboard interaktif berbasis web menggunakan framework Streamlit. Dashboard ini terdiri dari tiga halaman utama, yaitu halaman latar belakang aplikasi, halaman visualisasi berupa pie chart distribusi kelas, wordcloud, dan performa model, serta halaman analisis sentimen yang memungkinkan pengguna melakukan klasifikasi sentimen secara langsung melalui input teks.

REFERENSI

- [1] I. B. Permadi and A. Rokhman, “Implementasi Identitas Kependudukan Digital Dalam Upaya Pengamanan Data pribadi,” *JOPPAS: Journal of Public Policy and Administration Silampari*, vol. 4, no. 2, pp. 80–88, Jun. 2023, doi: 10.31539/joppas.v4i2.6199.
- [2] A. Luthfika Fairuz, R. Dias Ramadhani, N. Annisa, and F. Tanjung, “JURNAL DINDA Analisis Sentimen Masyarakat Terhadap COVID-19 Pada Media Sosial Twitter,” 2021. [Online]. Available: <http://jurnal.ittelkom-pwt.ac.id/index.php/dinda>
- [3] I. Maulana, W. Apriandari, and A. Pambudi, “ANALISIS SENTIMEN BERBASIS ASPEK TERHADAP ULASAN APLIKASI MYPERTAMINA MENGGUNAKAN SUPPORT VECTOR MACHINE,” 2023. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- [4] M. Izunnahdi, G. Aburrahman, and A. E. Wardoyo, “Sentimen Analisis Pada Data Ulasan Aplikasi KAI Access Di Google PlayStore Menggunakan Metode Multinomial Naive Bayes Sentiment Analysis on KAI Access Application Review Data on Google PlayStore Using Multinomial Naive Bayes Method,” 2021. [Online]. Available: <http://jurnal.unmuhammadiyah.ac.id/index.php/JST>
- [5] W. E. Saputro, H. Yuana, and W. D. Puspitasari, “ANALISIS SENTIMEN PENGGUNA DOMPET DIGITAL DANA PADA KOLOM KOMENTAR GOOGLE PLAY STORE DENGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE,” 2023.
- [6] R. A. Lestari, A. Erfina, and W. Jatmiko, “Penerapan Algoritma Support Vector Machine pada Analisis Sentimen Terhadap Identitas Kependudukan Digital,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 5, pp. 1063–1070, Oct. 2023, doi: 10.25126/jtiik.2023107264.
- [7] D. F. Setiawan, T. Tristiyanto, and A. Hijriani, “APLIKASI WEB SCRAPING DESKRIPSI PRODUK,” *Jurnal Teknoinfo*, vol. 14, no. 1, p. 41, Jan. 2020, doi: 10.33365/jti.v1i4i1.498.
- [8] Z. Alhaq, A. Mustopa, and J. D. Santoso, “PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER,” 2021.
- [9] D. Diandra Audiansyah, D. Eka Ratnawati, and B. Trias Hanggara, “Analisis Sentimen Aplikasi MyXL menggunakan Metode Support Vector Machine berdasarkan Ulasan Pengguna di Google Play Store,” 2022. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] A. M. Pravina, I. Cholissodin, and P. P. Adikara, “Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM),” 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] F. N. Dhewayani *et al.*, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-

DM,” *Jurnal Teknologi dan Informasi*, 2022, doi: 10.34010/jati.v12i1.

[12] G. Gumelar, Q. Ain, R. Marsuciati, S. Agustanti Bambang, A. Sunyoto, and M. Syukri Mustafa, “Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance,” 2021.

[13] F. Rassy, M. Kallista, and P. D. Wibawa, “Klasifikasi Kualitas Udara menggunakan Adaptive KNN dan Weighted KNN dengan Penggunaan SMOTE-Tomek Links dan Pendekatan Bagging,” 2024.

[14] C. Wulandari and L. Sunardi, “KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Aplikasi Spotify Pada Ulasan Pengguna di Google Play Store Menggunakan Metode Support Vector Machine,” *Media Online*, vol. 4, no. 5, pp. 2588–2595, 2024, doi: 10.30865/klik.v4i5.1762.

[15] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” 2021.