Machine Learning-Based Sentiment Analysis for Evaluating User Experience in the Access by KAI Application

Laili Nur Adiani*
Program Studi Akuntansi Sektor Publik, Politeknik Keuangan Negara STAN

Received: 17 September 2026 Accepted: 11 November 2026 Published: 1 December 2026

*4213250170_lailinuradiani@pknstan.ac.id

Kata Kunci: Analisis Sentimen, Access by KAI, Document Embedding, Support Vector Machine, Word Cloud

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak: Transformasi layanan pemesanan tiket digital melalui aplikasi Access by KAI menghasilkan berbagai reaksi pengguna di platform Google Play Store, yang didominasi oleh keluhan terkait stabilitas sistem. Penelitian ini bertujuan untuk mengevaluasi sentimen publik dan membandingkan kinerja empat algoritma klasifikasi (Support Vector Machine, Random Forest, Naive Bayes, dan k-Nearest Neighbors) dalam mendeteksi anomali aplikasi. Pendekatan metodologi diawali dengan ekstraksi 1000 data ulasan awal melalui web scraping, dilanjutkan dengan pelabelan ke dalam tiga kelas (positif, netral, dan negatif), serta penyeimbangan data menggunakan teknik random undersampling yang menghasilkan korpus akhir sebanyak 498 observasi data. Representasi teks menggunakan Document Embedding berbasis model fastText pra-latih berbahasa Indonesia (vektor 300 dimensi dengan agregasi rata-rata) diterapkan untuk menangkap konteks semantik bahasa, dan pengujian model dilakukan melalui 10-Fold Cross Validation. Hasil eksperimen menunjukkan bahwa algoritma Support Vector Machine (SVM) memperoleh tingkat akurasi tertinggi sebesar 80,5% dan nilai Area Under Curve (AUC) 0,935, secara signifikan mengungguli ketiga model lainnya. Analisis visual Word Cloud mengonfirmasi bahwa sentimen negatif secara spesifik berpusat pada kendala transaksi pembayaran dan akses aplikasi. Temuan ini memberikan kontribusi praktis berupa arsitektur sistem pemantauan sentimen otomatis berbasis SVM dan Document Embedding untuk membantu mengeksekusi perbaikan backend aplikasi berbasis data empiris.

PENDAHULUAN

Transformasi digital di sektor pelayanan publik Indonesia kini menjadi instrumen esensial, di mana PT Kereta Api Indonesia (Persero) terus melakukan inovasi layanan melalui pembaruan aplikasinya menjadi "*Access by KAI*". Pembaruan sistem ini dirancang untuk memperluas kapabilitas reservasi tiket elektronik dan integrasi layanan antarmoda secara terpusat bagi masyarakat. Namun, dalam ekosistem bisnis digital yang sangat dinamis, stabilitas perangkat lunak dan kualitas pengalaman pengguna (*user experience*) memiliki korelasi langsung dengan citra korporat dan tingkat kepercayaan publik. Penelitian sebelumnya menunjukkan bahwa kualitas aplikasi *mobile* sangat dipengaruhi oleh keandalan fungsionalnya, karena kesalahan fungsional dan aplikasi yang berhenti bekerja (*crash*) merupakan keluhan yang paling sering disampaikan pengguna dalam ulasan [1]. Oleh karena itu, PT KAI memiliki urgensi bisnis untuk secara proaktif memantau dan mengevaluasi kepuasan penumpang agar platform penjualan tiket digital mereka dapat beroperasi secara optimal dan berkelanjutan.

Meskipun pembaruan aplikasi ditujukan untuk meningkatkan layanan, realita interaksi digital pengguna memunculkan tantangan teknis. Observasi pada repositori publik di *Google Play Store* menunjukkan tingginya gelombang ulasan yang didominasi oleh sentimen negatif pasca-pembaruan

sistem Access by KAI. Pengguna secara spesifik mengeluhkan hambatan fungsionalitas, seperti kegagalan transaksi pembayaran, sistem antarmuka yang sering tidak merespons (*bug/error*), hingga kendala dalam memuat jadwal kereta api. Penurunan kepuasan pengguna (*user satisfaction*) yang tidak termitigasi dengan baik berisiko memicu hilangnya loyalitas pelanggan terhadap ekosistem digital perusahaan, karena kepuasan pelanggan terbukti berpengaruh terhadap loyalitas pelanggan pada layanan elektronik [2]. Ulasan pengguna pada platform digital merupakan sumber informasi penting yang dapat dimanfaatkan perusahaan untuk mengevaluasi kualitas layanan dan memahami persepsi pelanggan secara lebih mendalam.

Namun, jumlah ulasan yang sangat besar membuat proses analisis manual menjadi tidak efektif dan memerlukan waktu yang lama. Oleh karena itu, diperlukan pendekatan otomatis melalui analisis sentimen (*sentiment analysis*) untuk mengelompokkan opini pengguna ke dalam kategori sentimen positif, negatif, maupun netral. Analisis sentimen merupakan bagian dari *text mining* yang bertujuan mengidentifikasi kecenderungan opini masyarakat terhadap suatu produk atau layanan berdasarkan data tekstual [3]. Mengingat metrik kuantitatif sederhana berupa rating bintang tidak mampu memvisualisasikan akar permasalahan teknis secara utuh, diperlukan sebuah instrumen analitik tingkat lanjut. Sebagai solusi analitik atas permasalahan tersebut, penelitian ini mengimplementasikan analisis sentimen berbasis *text mining*. Untuk mengatasi kelemahan metode tradisional seperti TF-IDF yang sering kali menghilangkan struktur dan konteks semantik kata, penelitian ini mengadopsi teknik *Document Embedding*. Teknik ini bekerja dengan memproyeksikan makna semantik ulasan pengguna ke dalam representasi vektor berdimensi padat (*dense vector*), sehingga pola kata yang berkaitan dengan keluhan dapat terdeteksi secara lebih akurat [4].

Dalam beberapa tahun terakhir, penerapan machine learning pada analisis sentimen mengalami perkembangan yang signifikan. Berbagai algoritma klasifikasi seperti *Support Vector Machine (SVM), Naive Bayes, K-Nearest Neighbor (KNN),* dan *Random Forest* banyak digunakan karena mampu menghasilkan performa klasifikasi yang baik pada data teks. Penelitian sebelumnya menunjukkan bahwa algoritma *Random Forest* mampu menghasilkan akurasi sebesar 84% dalam analisis sentimen ulasan aplikasi Dana pada *Google Play Store [3]*. Selain itu, penelitian Devi et al. [5] mengenai analisis sentimen aplikasi *Access by KAI* menggunakan metode Word2Vec dan SVM menghasilkan akurasi sebesar 81% dengan kernel RBF. Penelitian lain oleh Wirayudha et al. [6] menggunakan metode IndoBERT juga menunjukkan efektivitas deep learning dalam klasifikasi sentimen ulasan aplikasi *Access by KAI* dengan akurasi sebesar 76%. Meskipun demikian, penelitian terdahulu pada ulasan *Access by KAI* [5], [6] masing-masing hanya menguji satu pendekatan (*Word2Vec* dengan SVM dan IndoBERT). Berdasarkan penelusuran penulis, belum ditemukan penelitian yang membandingkan secara bersamaan kinerja empat algoritma *machine learning* (SVM, *Random Forest*, *Naive Bayes*, dan kNN) dalam mengklasifikasikan sentimen pengguna *Access by KAI* dengan representasi *Document Embedding* menggunakan platform visual *data mining* seperti *Orange Data Mining*. *Orange Data Mining* dipilih karena menyediakan antarmuka visual yang memudahkan proses *preprocessing, feature extraction,* pemodelan, hingga evaluasi performa tanpa memerlukan pemrograman yang kompleks. Selain itu, penggunaan *document embedding* berbasis *fastText* pada Orange memungkinkan representasi teks menjadi vektor numerik sehingga dapat meningkatkan kualitas klasifikasi sentimen.

Penelitian menggunakan data sebanyak 1000 ulasan pengguna *Access by KAI* yang diperoleh melalui teknik *web scraping* menggunakan *google play scraper* pada Google Play Store. Kriteria ulasan berbahasa Indonesia dan minimal terdiri atas lima kata agar informasi yang diperoleh lebih spesifik untuk dianalisis. Sebelum dilakukan proses klasifikasi, data mengalami tahap *balancing* karena distribusi awal data tidak seimbang, yaitu terdiri atas 166 sentimen positif, 561 sentimen negatif, dan 273 sentimen netral. Penelitian ini bertujuan untuk mengevaluasi sentimen publik terhadap fungsionalitas aplikasi Access by KAI pasca pembaruan, sekaligus membandingkan performa empat algoritma klasifikasi yakni *Support Vector Machine* (SVM), *Random Forest*, *Naive Bayes*, dan *k-Nearest Neighbors* (kNN) pada 498 set data ulasan yang telah diseimbangkan (*balanced data*). Melalui riset ini, diharapkan PT KAI dapat mengadopsi model algoritma terbaik sebagai sistem pemantauan sentimen keluhan secara otomatis, sehingga kebijakan perbaikan *backend* aplikasi dapat dieksekusi secara presisi dan berbasis data (*data driven*).

TINJAUAN LITERATUR

Data Mining

Data mining merupakan proses komputasional analitik yang dirancang untuk mengeksplorasi dan mengekstraksi informasi, pola, atau anomali yang tersembunyi dari himpunan data berukuran masif *(big data*). Dalam perkembangannya, ketika objek data yang ditambang berupa bahasa alami (*Natural Language Processing*) yang tidak terstruktur seperti dokumen atau teks, disiplin ilmu ini diadaptasi menjadi *text mining* (penambangan teks). Teknik komputasi ini sangat krusial dalam rekayasa perangkat lunak dan intelijen bisnis karena memungkinkan pengambilan keputusan strategis yang sepenuhnya didasarkan pada rekam jejak data empiris (*data driven*) [7].

Analisis Sentimen

Analisis sentimen adalah salah satu cabang implementasi paling signifikan dari *text mining*. Metode ini merupakan instrumen komputasional yang berfungsi untuk mengidentifikasi, mengekstraksi, dan memetakan polaritas opini publik yang bersifat kualitatif menjadi metrik analitik kuantitatif yang terukur. Melalui analisis sentimen, ulasan yang subjektif dapat diklasifikasikan secara otomatis ke dalam kategori sentimen positif, netral, atau negatif [8]. Dalam evaluasi layanan digital, metode ini terbukti tangguh sebagai alat prediktif untuk mendeteksi arah preferensi, keluhan, dan tingkat kepuasan konsumen.

Web Scraping

Web scraping adalah sebuah teknik komputasi otomatis yang diimplementasikan untuk mengekstraksi data dalam jumlah besar dari situs web atau antarmuka aplikasi. Mengingat kapasitas data ulasan di repositori publik dapat mencapai ribuan entri, pengumpulan data secara manual menjadi tidak efisien dan rentan kesalahan. Dengan menggunakan *script web scraping* melalui Antarmuka Pemrograman Aplikasi (API), peneliti dapat menarik berbagai atribut metadata pengguna secara serentak seperti teks ulasan, stempel waktu (*timestamp*), dan peringkat bintang (*rating*) yang kemudian digunakan sebagai bahan baku fundamental untuk penelitian [3].

Text Preprocessing (Prapemrosesan Teks)

Text preprocessing adalah tahapan standardisasi linguistik yang wajib dilakukan sebelum data teks mentah dimasukkan ke dalam model komputasi. Data yang bersumber dari ulasan publik umumnya dipenuhi *noise* seperti kesalahan ejaan, tanda baca, tautan (URL), serta kata-kata tidak baku. Kusuma dan Cahyono [9] menjelaskan bahwa agar algoritma dapat memetakan pola bahasa secara presisi, teks harus melewati pembersihan yang meliputi: *case folding* (penyeragaman ke huruf kecil), *cleansing* (eliminasi karakter non-teks), *tokenization* (segmentasi kalimat menjadi kata tunggal), dan *stopword removal* (penghapusan kata hubung yang tidak memiliki muatan polaritas sentimen).

Balancing Data (Penyeimbangan Data)

Ketidakseimbangan data (*class imbalance*) merupakan fenomena di mana salah satu kategori sentimen (misalnya ulasan negatif) memiliki jumlah yang jauh lebih dominan dibandingkan kategori lainnya. Wibowo et al. [10] menegaskan bahwa melatih model *machine learning* menggunakan data asimetris akan menyebabkan model mengalami *overfitting* atau bias terhadap kelas mayoritas. Untuk mengatasi hal ini, teknik *balancing data* seperti *undersampling* (mereduksi kelas mayoritas secara acak) atau *oversampling* (menggandakan data minoritas) harus diterapkan guna memastikan algoritma belajar mengenali seluruh kelas secara objektif dan setara [11].

Document Embedding

Document Embedding adalah metodologi representasi teks mutakhir yang memproyeksikan struktur dokumen teks ke dalam ruang vektor berdimensi rendah yang sangat padat (*dense continuous vectors*). Pendekatan ini merupakan penyempurnaan dari metode tradisional seperti *Term Frequency-Inverse Document Frequency* (TF-IDF) yang menghasilkan matriks renggang (*sparse matrix)* dan mengabaikan urutan kata. Berlandaskan riset Le dan Mikolov [4], arsitektur embedding mampu membedah dependensi semantik antar-kata secara utuh. Melalui teknik ini, algoritma dapat mengenali bahwa leksikon seperti "gagal", "bug", dan "error" memiliki kedekatan jarak spasial sebagai satu kesatuan representasi dari keluhan teknis. Perlu dicatat bahwa *Doc2Vec* (*Paragraph Vector*) yang diusulkan Le dan Mikolov [4] hanyalah salah satu varian *Document Embedding*. Pada penelitian ini, representasi dokumen dihasilkan oleh widget *Document Embedding* pada *Orange Data Mining* yang menggunakan model *fastText* pra-latih[12]. Setiap

token pada ulasan dipetakan ke vektor berdimensi 300 menggunakan model berbahasa Indonesia, kemudian vektor seluruh token dalam satu ulasan diagregasi dengan rata-rata (*mean*) menjadi satu vektor dokumen berdimensi 300. Token di luar kosakata ditangani melalui penjumlahan vektor *n-gram* karakter penyusunnya sehingga kata tidak baku pada ulasan tetap memperoleh representasi vektor. *Doc2Vec* tidak digunakan.

Algoritma *Machine Learning*

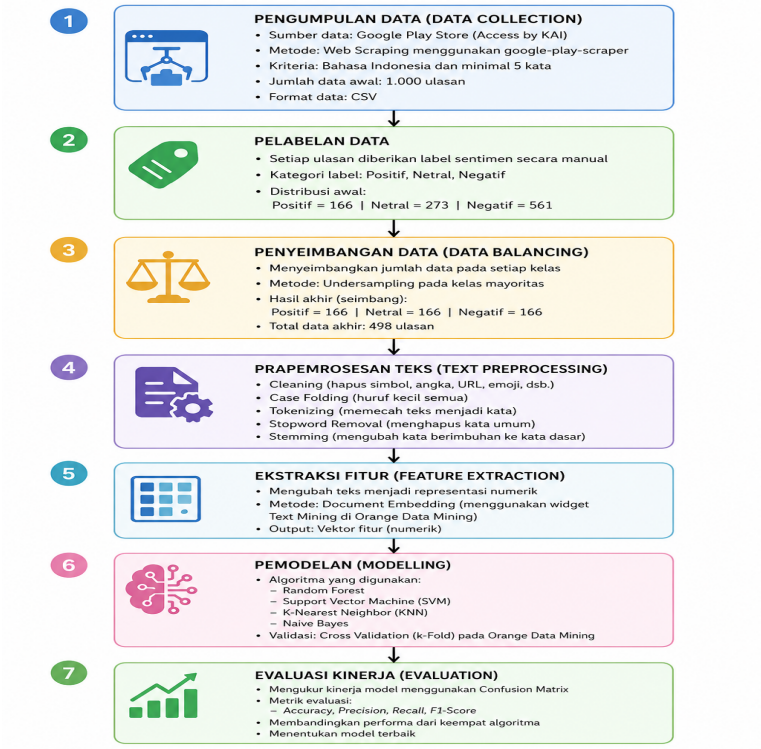
Machine learning adalah sub-bidang kecerdasan buatan yang memungkinkan sistem untuk belajar mengenali pola data tanpa harus diprogram secara eksplisit. Dalam klasifikasi sentimen, penelitian ini membandingkan empat algoritma *supervised learning* (pembelajaran terarah) untuk menentukan model terbaik di antara *Support Vector Machine (SVM)* [13] , *Random Forest* [14], *Naive Bayes* [15], dan *k-Nearest Neighbors (kNN)* [16]. SVM juga telah terbukti efektif untuk kategorisasi teks yang berdimensi tinggi [17].

Access by KAI

Access by KAI adalah platform digital yang menjadi garda operasional dari transformasi pelayanan elektronik PT Kereta Api Indonesia (Persero). Aplikasi ini dirancang untuk mengintegrasikan layanan reservasi e-tiket dan manajemen mobilitas antarmoda secara nasional. Migrasi sistem pemesanan dan pembaruan antarmuka pada layanan publik ini memicu pergeseran pengalaman (*user experience*) yang terekam secara komprehensif pada repositori ulasan Google Play Store, sehingga menjadikannya objek penelitian yang ideal untuk evaluasi kepuasan layanan publik berbasis *machine learning*.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif komputasional untuk melakukan analisis sentimen berbasis *Natural Language Processing* (NLP). Seluruh siklus pemrosesan data, mulai dari prapemrosesan hingga komparasi algoritma *machine learning*, dieksekusi secara visual menggunakan perangkat lunak *Orange Data Mining*. Alur penelitian secara komprehensif dirancang ke dalam lima tahapan utama sebagaimana diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pengumpulan Data (*Data Collection*)

Populasi bersumber dari ulasan pengguna Access by KAI di Google Play Store. Sampel ditarik menggunakan metode *web scraping* melalui pustaka Python (*google play scraper*). Data dibatasi pada 1.000 ulasan terbaru yang telah disaring dari ulasan sangat pendek (minimal 5 kata). Proses ini menggunakan teknik *crawling* otomatis untuk mendapatkan gambaran persepsi pengguna terkini.

Tabel 1. Pengumpulan Data

Komponen	Keterangan
Sumber Data	*Google Play Store (Aplikasi Access by KAI)*
Teknik	*Web Scraping* menggunakan library *google-play-scraper*
Volume Data	1.000 ulasan (Dataset Terfilter)
Atribut	Nama Pengguna, *Rating,* Tanggal, Teks Ulasan

Sumber: diolah penulis

Pelabelan Data

Anotasi sentimen ke dalam tiga kelas (Positif, Netral, Negatif). Evaluasi sentimen didasarkan pada pembacaan substansi komentar secara tekstual, bukan sekadar bersandar pada rating bintang. Hal ini krusial untuk mengoreksi anomali ketidaksesuaian antara rating dan makna sebenarnya dari ulasan pengguna. Pelabelan dilakukan secara manual oleh satu annotator (peneliti) dengan berpedoman pada kriteria pada Tabel 2. Karena hanya melibatkan satu annotator, tidak terdapat disagreement antar-annotator yang perlu diselesaikan dan inter-annotator agreement (Cohen's Kappa) tidak dapat dihitung. Pelabelan manual berpotensi memuat subjektivitas sehingga kondisi ini diakui sebagai keterbatasan penelitian.

Tabel 2. Pelabelan Data

Kriteria Label	Keterangan
Sentimen Positif	Ulasan berisi apresiasi atau kepuasan terhadap fitur aplikasi.
Sentimen Netral	Ulasan berisi saran atau komentar objektif tanpa muatan emosi kuat.
Sentimen Negatif	Ulasan berisi keluhan, laporan *bug*, atau ketidakpuasan layanan.
Koreksi Manual	Menyelaraskan teks ulasan dengan label meskipun *rating* bintang tidak sesuai (misal: Bintang 5 namun teks berisi komplain berat).

Sumber: diolah penulis

Penyeimbangan Data (*Data Balancing*)

Fenomena ketidakseimbangan kelas (*class imbalance*) sering terjadi pada data publik komersial, di mana sentimen negatif kerap mendominasi. Jika algoritma dilatih menggunakan data yang timpang, model akan cenderung mengalami *overfitting* atau bias prediksi terhadap kelas mayoritas [10] . Berdasarkan hasil pelabelan awal, jumlah ulasan positif merupakan kelas minoritas dengan total 166 ulasan yang tervalidasi. Untuk memastikan model belajar mengenali pola secara setara, penelitian ini menerapkan teknik random *undersampling*. Data ulasan pada kelas negatif dan netral masing-masing direduksi secara acak hingga menyamai jumlah kelas minoritas (166 ulasan). Melalui prosedur ini, terbentuklah korpus data latih (*training dataset*) baru berjumlah 498 ulasan yang seimbang secara proporsional (tepat 33,3% untuk setiap kelas sentimen)

Prapemrosesan Teks (Text Preprocessing)

Teks ulasan mentah dibersihkan agar algoritma dapat memproses kata kunci secara efektif. Proses ini dilakukan melalui widget *Preprocess Text* pada *Orange Data Mining*.

Tabel 3.Text Preprocessing

Tahap Preprocessing	Fungsi Proses
Case Folding	Menyeragamkan seluruh karakter teks menjadi huruf kecil (*lowercase*).
Cleansing	Menghapus elemen non-tekstual menggunakan *Regular Expression* (RegEx), termasuk eliminasi angka, tanda baca, simbol, emotikon, dan tautan (URL).
Tokenization	Mengurai kalimat yang utuh menjadi unit-unit kata dasar (*token*).
Stopword Removal	Mengeliminasi kata-kata umum dalam bahasa Indonesia (seperti "dan", "yang", "di", "ke") yang sering muncul namun tidak memiliki signifikansi terhadap pembobotan polaritas sentimen.

Sumber: diolah penulis

Ekstraksi Fitur *(Feature Extraction)*

Penelitian meninggalkan metode ekstraksi fitur konvensional yang berbasis frekuensi kemunculan kata (Bag-of-Words seperti TF-IDF) dan beralih menggunakan pendekatan mutakhir, yakni Document Embedding.

Tabel 4. Ekstraksi Fitur

Widget	Keterangan
Document Embedding	Model *embedding* memproyeksikan seluruh struktur dokumen teks ke dalam ruang semantik komputasional [4]. Metode ini memungkinkan sistem untuk mempelajari korelasi sintaksis dan dependensi antar kata. Sebagai dampaknya, algoritma mampu mengenali berbagai variasi kata keluhan (misalnya: "gagal", "bug", "error", "macet") sebagai satu kesatuan konteks permasalahan teknis, yang sangat krusial dalam meningkatkan akurasi deteksi sentimen negatif.

Sumber: diolah penulis

Pemodelan (*Modelling*)

Penelitian membandingkan empat algoritma *machine learning* untuk menemukan model klasifikasi terbaik yang paling sesuai dengan karakteristik ulasan Access by KAI.

Tabel 5. Model Algoritma

Algoritma	Keterangan
Support Vector Machine (SVM)	Memisahkan data dengan membangun hiperbidang (*hyperplane*) optimal untuk mencari margin pemisah maksimum antar kelas [13].
Naive Bayes (NB)	Pengklasifikasi probabilistik cepat berbasis Teorema Bayes dengan asumsi independensi fitur bersyarat [18].
Random Forest	Algoritma ensemble yang memitigasi varians prediksi melalui agregasi hasil dari sejumlah besar pohon keputusan (decision tree) yang dibangun secara independen [14].
k-Nearest Neighbors (kNN)	Metode pemodelan berbasis jarak (umumnya jarak *Euclidean*) yang menentukan sentimen dokumen berdasarkan kedekatan spasial dengan *k* tetangga sampel latihnya [9], [16].

Sumber: diolah penulis

Seluruh algoritma dijalankan menggunakan pengaturan bawaan (*default*) pada widget model di *Orange Data Mining* tanpa penyetelan hiperparameter, dengan parameter sebagai berikut: SVM menggunakan kernel RBF dengan nilai *cost* C = 1,0 dan toleransi numerik 0,001; *Random Forest* menggunakan 10 pohon keputusan; kNN menggunakan k = 5 tetangga dengan jarak *Euclidean* dan pembobotan seragam; sedangkan *Naive Bayes* tidak memiliki parameter yang dapat diatur pada widget Orange sehingga tidak terdapat nilai alpha yang disetel.

Evaluasi Kinerja (*Evaluation*)

Tahap akhir adalah mengukur sejauh mana model mampu memprediksi sentimen dengan benar menggunakan metrik standar. Untuk menjamin objektivitas pengujian dan mencegah bias akibat partisi data yang statis, performa keempat algoritma divalidasi menggunakan teknik *10-Fold Cross Validation [19]*. Metode ini membagi 498 data menjadi 10 lipatan bagian yang sama besar. Kinerja model kemudian diekstraksi dari hasil *Confusion Matrix* dan dikuantifikasi menggunakan metrik standar evaluasi sistem pakar, yang mencakup nilai *Area Under ROC Curve* (AUC) [20], *Classification Accuracy* (CA), F1-Score, *Precision* (Presisi), dan *Recall* (Sensitivitas). Model dengan nilai metrik akumulatif tertinggi akan diusulkan sebagai instrumen analitik sentimen terbaik untuk diterapkan oleh pengembang Access by KAI.

Tabel 6. Metrik Evaluasi

Metrik Evaluasi	Definisi
Accuracy	Rasio prediksi benar dibandingkan dengan total keseluruhan data.
Precision	Tingkat ketepatan antara informasi yang diminta dengan jawaban yang diberikan oleh sistem.
Recall	Tingkat keberhasilan sistem dalam menemukan kembali informasi.
F1-Score	Nilai rata-rata harmonis yang menyeimbangkan antara *Precision* dan *Recall*.
Validation Method	*10-Fold Cross Validation [19]* (Membagi data menjadi 10 bagian untuk diuji secara bergantian).

Sumber: diolah penulis

HASIL PENELITIAN DAN PEMBAHASAN

Bagian ini memaparkan temuan empiris dari hasil pemrosesan data menggunakan Orange Data Mining. Analisis difokuskan pada perbandingan performa algoritma klasifikasi dan interpretasi sentimen pengguna terhadap aplikasi Access by KAI.

Distribusi Data Sentimen

Berdasarkan proses pelabelan terhadap 1.000 ulasan yang telah melewati tahap penyaringan, ditemukan pola persepsi pengguna yang didominasi oleh ketidakpuasan. Distribusi polaritas sentimen disajikan secara rinci pada Tabel 7.

Tabel 7. Distribusi Data Sentimen

Kelas Sentimen	*Jumlah Ulasan*	*Persentase*
Negatif	*561*	*56,1%*
Netral	*273*	*27,3%*
Positif	*166*	*16,6%*
TOTAL	*1000*	*100%*

Sumber: diolah penulis

Tabel 7 memperlihatkan bahwa lebih dari 50% pengguna memberikan respons negatif. Hal ini berbanding lurus dengan tren ulasan di Google Play Store pasca-pembaruan sistem, di mana aspek stabilitas aplikasi menjadi keluhan utama.

Distribusi Data Sentimen setelah Penyeimbangan Data (*Data Balancing*)

Ketidakseimbangan kelas sering kali menjadi bias utama dalam klasifikasi sentimen ulasan publik. Penerapan random *undersampling* telah berhasil mengumpulkan 498 data latih menjadi proporsi yang simetris (166 ulasan untuk masing-masing kelas Positif, Netral, dan Negatif). Penggunaan korpus yang seimbang ini menjamin bahwa model belajar mengenali pola semantik bahasa secara objektif, bukan sekadar memprediksi probabilitas kelas yang paling sering muncul.

Tabel 8. Distribusi Data Sentimen setelah *Data Balancing*

Sentimen	*Jumlah Sebelum Data Balancing*	*Jumlah Setelah Data Balancing*	*Persentase Setelah Data Balancing*
Negatif	*561*	*166*	*33,33%*
Netral	*273*	*166*	*33,33%*
Positif	*166*	*166*	*33,33%*
TOTAL	*1000*	*498*	*100%*

Sumber: diolah penulis

Perbandingan Performa Model Klasifikasi

Pengujian keandalan model dilakukan menggunakan metode *10-Fold Cross Validation*. Metrik evaluasi mencakup *Area Under ROC Curve (AUC), Classification Accuracy (CA), F1-Score, Precision*, dan *Recall*. Hasil pengujian keempat model dirangkum dalam Tabel 9.

Tabel 9. Hasil Performa Model Klasifikasi

No	*Model*	*AUC*	*CA*	*F1-Score*	*Precision*	*Recall*
1	*Support Vector Machine (SVM)*	0,935	0,805	0,806	0,808	0,805
2	*Random Forest*	0,902	0,741	0,743	0,745	0,741
3	*Naive Bayes (NB)*	0,892	0,739	0,739	0,739	0,739
4	*k-Nearest Neighbors (kNN)*	0,836	0,673	0,670	0,673	0,673
(Catatan: Nilai disajikan dalam rentang 0-1 dan dibulatkan menjadi tiga angka di belakang koma).						

Sumber: diolah penulis dari orange data mining

Analisis Komparatif Performa Algoritma

Integrasi teknik Document Embedding memberikan dampak yang sangat terukur terhadap kinerja klasifikasi keempat model. Berdasarkan Tabel 9, temuan komparatif dapat diuraikan sebagai berikut:

1. *Support Vector Machine (SVM)*: Model SVM mendominasi performa dengan mencatatkan nilai tertinggi secara absolut di seluruh metrik. SVM meraih tingkat akurasi (CA) sebesar 0,805 (80,5%) dan presisi 0,808. Nilai AUC sebesar 0,935 merepresentasikan bahwa SVM memiliki tingkat pemisahan kelas (*class separability)* pada kategori *excellent* (AUC ≥ 0,9) menurut pedoman interpretasi nilai AUC [21]. Kinerja superior ini membuktikan bahwa algoritma SVM memiliki kompatibilitas yang sangat tinggi terhadap matriks vektor padat (*dense vectors*) yang dihasilkan oleh *Document Embedding*. SVM mampu menemukan hiperbidang linier (*hyperplane*) paling optimal untuk menarik batas tegas antara leksikon keluhan, saran teknis, dan pujian pengguna.
2. *Random Forest* dan *Naive Bayes*: *Random Forest* menempati posisi kedua dengan tingkat akurasi 0,741 (74,1%), diikuti oleh Naive Bayes dengan akurasi 0,739 (73,9%). Pendekatan *ensemble learning* pada Random Forest cukup tangguh dalam mengagregasi prediksi dari banyak pohon keputusan, sehingga kebal terhadap *noise* teks. Di sisi lain, capaian *Naive Bayes* dengan probabilitas bersyaratnya terbilang sangat efisien; meskipun menggunakan komputasi yang jauh lebih ringan, model ini mampu mengimbangi kompleksitas Random Forest.
3. *k-Nearest Neighbors (kNN):* Algoritma kNN menghasilkan akurasi terendah, yakni 0,673 (67,3%). Meskipun nilainya telah melampaui probabilitas acak, pendekatan berbasis spasial pada kNN masih kurang optimal. Penggunaan perhitungan jarak cenderung kehilangan efektivitasnya dalam mendefinisikan batas keputusan (*decision boundary*) yang kompleks di dalam ruang dimensi vektor embedding, jika dibandingkan dengan algoritma optimasi matematis seperti SVM.

Analisis *Confusion Matrix Model SVM*

Untuk menganalisis lebih dalam bagaimana model *Support Vector Machine* (SVM) menghasilkan akurasi tertinggi sebesar 80,5%, dilakukan analisis evaluasi terhadap *Confusion Matrix*. Matriks konfusi ini memetakan perbandingan antara label aktual (*ground truth*) dengan label prediksi yang dihasilkan oleh mesin pada 498 data latih yang seimbang. Hasil matriks konfusi untuk model SVM disajikan pada Tabel 10. Setiap sel berisi jumlah ulasan dan diagonal utama menunjukkan prediksi yang benar; total prediksi benar adalah 143 + 131 + 127 = 401 dari 498 ulasan (80,5%), sesuai dengan nilai *Classification Accuracy* (CA) pada Tabel 9.

Tabel 10. Confusion Matrix Model Support Vector Machine (SVM)

	Prediksi: Positif	Prediksi: Netral	Prediksi: Negatif	Total Aktual
Aktual: Positif	**143**	0	23	166
Aktual: Netral	14	**131**	21	166
Aktual: Negatif	19	20	**127**	166

Sumber: diolah penulis dari orange data mining

Berdasarkan Tabel 10, perilaku klasifikasi model SVM dapat disimpulkan ke dalam tiga poin utama:

1. Kelas Positif: Model berhasil menebak dengan tepat 143 dari 166 ulasan (86,1%). Hal ini membuktikan bahwa leksikon kepuasan (seperti "bagus", "mudah") memiliki fitur semantik yang sangat khas sehingga mudah dipisahkan oleh *hyperplane* SVM.
2. Kelas Netral: Prediksi benar mencapai 131 ulasan (78,9%). Sebagian kecil kesalahan klasifikasi mayoritas jatuh ke prediksi Negatif (21 ulasan), yang disebabkan oleh kemiripan kosakata antara saran perbaikan aplikasi (Netral) dengan keluhan murni (Negatif).
3. Kelas Negatif: Model memprediksi tepat 127 ulasan (76,5%). Kesalahan tebak ke arah kelas Positif (19 ulasan) umumnya dipicu oleh penggunaan gaya bahasa sarkasme pengguna (misal: menggunakan kata "hebat" untuk mengkritik aplikasi error, "Hebat sekali aplikasinya, saking hebatnya sampai tidak bisa login"), yang mengecoh batas keputusan linear algoritma.

Secara keseluruhan, sebaran prediksi benar yang merata di atas 76% untuk ketiga kelas membuktikan bahwa model SVM tidak bias dan sangat stabil dalam mengklasifikasikan dokumen ulasan pengguna.

Perbandingan dengan Penelitian Terdahulu

Untuk menempatkan hasil penelitian ini dalam konteks penelitian sebelumnya, Tabel 11 membandingkan akurasi model terbaik pada penelitian ini dengan penelitian terdahulu yang dirujuk pada bagian Pendahuluan.

Tabel 11. Perbandingan Akurasi dengan Penelitian Terdahulu

Penelitian	Objek Ulasan	Metode	Akurasi
Larasati et al. [3]	Aplikasi Dana	*Random Forest*	84%
Devi et al. [5]	*Access by KAI*	*Word2Vec* + SVM (kernel RBF)	81%
Wirayudha et al. [6]	*Access by KAI*	IndoBERT	76%
Penelitian ini	*Access by KAI*	*Document Embedding* (*fastText*) + SVM (terbaik dari empat algoritma)	80,5%

Sumber: diolah penulis dari [3], [5], [6] dan Orange Data Mining

Akurasi SVM pada penelitian ini (80,5%) hampir setara dengan hasil Devi et al. [5] (81%) pada objek yang sama dan lebih tinggi daripada hasil Wirayudha et al. [6] (76%) yang menggunakan IndoBERT, sedangkan hasil Larasati et al. [3] (84%) diperoleh pada aplikasi yang berbeda. Perbandingan ini bersifat indikatif karena perbedaan dataset, periode pengambilan data, jumlah dan komposisi kelas, serta skema pengujian antar-penelitian. Kontribusi utama penelitian ini terletak pada perbandingan empat algoritma pada representasi *Document Embedding* dan skema *10-Fold Cross Validation* yang sama, sehingga perbedaan kinerja antar-algoritma dapat diamati secara berdampingan.

Analisis Topik dan Visualisasi *Word Cloud*

Analisis kualitatif dilakukan melalui visualisasi *Word Cloud*. Ekstraksi ini memetakan kata-kata dengan frekuensi kemunculan (*Term Frequency*) tertinggi pada masing-masing polaritas kelas. Ukuran teks pada *Word Cloud* merepresentasikan tingkat dominasi kata tersebut di dalam korpus ulasan pengguna Access by KAI.

1. Analisis Sentimen Positif

Pada visualisasi *Word Cloud* kelas Positif, leksikon yang paling mendominasi adalah kata "mudah", "cepat", "bagus", "nyaman", dan "mantap". Kemunculan kata-kata ini mengindikasikan bahwa sebagian pengguna sangat mengapresiasi efisiensi aplikasi. Kinerja utama yang memuaskan pengguna berpusat pada kemudahan proses transaksi (kata "beli", "pesan", "bayar"). Kosakata pujian pada kelas ini memiliki struktur semantik yang sangat khas dan eksplisit. Hal ini secara empiris menjawab mengapa model SVM pada matriks konfusi sebelumnya mampu mendeteksi kelas positif dengan sensitivitas paling tinggi (mencapai 86,1%), karena fitur kata-katanya sangat terpisah dari kelas komplain.

Gambar 2. Word Cloud Positif



Sumber: diolah penulis dari orange data mining

2. Analisis Sentimen Netral

Distribusi kata pada kelas netral didominasi oleh kata benda dan kata kerja operasional tanpa muatan emosi yang kuat, seperti "jadwal", "tiket", "beli", "bayar", dan "aplikasi". Analisis topik pada kelas ini menunjukkan bahwa ulasan netral umumnya berisi pernyataan objektif, pertanyaan informatif seputar cara mengecek jadwal kereta, atau saran pengembangan fitur. Menariknya, terdapat irisan (overlap) kosakata fungsional (seperti "bayar" dan "tiket") yang juga sering muncul di kelas negatif. Irisan leksikon inilah yang secara komputasional menjadi tantangan bagi algoritma machine learning, sehingga menyebabkan beberapa ulasan netral di matriks konfusi rentan terklasifikasi secara keliru menjadi kelas negatif.

Gambar 3. Word Cloud Netral



Sumber: diolah penulis dari orange data mining

3. Analisis Sentimen Negatif

Visualisasi kelas memberikan gambaran yang sangat jelas mengenai titik kelemahan sistem aplikasi Access by KAI pasca pembaruan. Kata kunci yang muncul sangat masif adalah "error", "gagal", "susah", "buka", "update", dan "kecewa". Narasi komplain pengguna terpusat pada kendala teknis (*sistem bug*), di mana aplikasi menjadi sulit diakses ("susah", "buka", "lambat") atau terjadinya kegagalan transaksi setelah aplikasi diperbarui ("update", "bayar", "gagal"). Kumpulan kosakata ini memvalidasi hipotesis awal bahwa lonjakan sentimen negatif di Google Play Store bukan disebabkan oleh fitur yang kurang lengkap, melainkan karena stabilitas *backend* dan performa antarmuka sistem yang sangat krusial mengganggu kelancaran pengguna dalam membeli tiket.

Gambar 4. Word Cloud Negatif



Sumber: diolah penulis dari orange data mining

Implikasi Temuan terhadap Bisnis Access by KAI

Secara komprehensif, validasi akurasi 80,5% dari model SVM yang selaras dengan temuan kualitatif pada *Word Cloud* menegaskan bahwa opini publik adalah aset data yang valid. Tingginya frekuensi kata "error" dan "gagal" pada sentimen negatif membuktikan bahwa keluhan pengguna adalah masalah teknis yang nyata di dalam aplikasi. Secara manajerial, arsitektur *text mining* (berbasis *Document Embedding* dan SVM) yang dilakukan dalam penelitian ini dapat diadopsi sebagai *Automated Monitoring System* di divisi IT PT Kereta Api Indonesia (Persero). Melalui sistem pemantauan otomatis ini, pihak manajemen tidak perlu lagi membaca ribuan ulasan secara manual; sistem akan langsung mengekstraksi dan menyortir keluhan spesifik secara *real-time*. Selain itu, pengembang PT KAI dapat segera mengetahui letak kerusakan pada aplikasi. Hasilnya, proses perbaikan server, penanganan sistem yang bermasalah (bug), dan peningkatan kualitas layanan dapat dilakukan dengan jauh lebih cepat dan tepat sasaran. Respon yang cepat dari perusahaan akan menjaga kepercayaan dan kenyamanan masyarakat dalam menggunakan aplikasi Access by KAI.

KESIMPULAN

Berdasarkan hasil penelitian dan pengujian model komputasional, dapat disimpulkan bahwa integrasi ekstraksi fitur *Document Embedding* dengan algoritma *Support Vector Machine* (SVM) merupakan model terbaik untuk mengevaluasi sentimen pengguna aplikasi Access by KAI. Model SVM berhasil mengungguli algoritma Random Forest, Naive Bayes, dan k-Nearest Neighbors (kNN) dengan tingkat akurasi tertinggi sebesar 80,5% dan nilai Area Under Curve (AUC) 0,935 pada 498 data latih yang telah diseimbangkan. Keberhasilan model ini dalam memetakan polaritas opini publik secara akurat juga mengindikasikan fakta bahwa sentimen negatif pengguna bersumber pada kelemahan sistem. Melalui ekstraksi leksikon ulasan, keluhan terbukti sangat didominasi oleh masalah teknis seperti kegagalan transaksi pembayaran, kendala login, dan aplikasi yang sering tidak merespons (error/bug) pasca-pembaruan aplikasi, yang secara langsung menurunkan tingkat kepuasan pelanggan digital PT Kereta Api Indonesia (Persero).

Penelitian ini memiliki beberapa keterbatasan. Pertama, data hanya bersumber dari satu platform, yaitu *Google Play Store*, sehingga hasilnya belum tentu merepresentasikan opini pengguna di platform lain. Kedua, pelabelan dilakukan secara manual oleh satu *annotator* sehingga berpotensi subjektif dan tingkat kesepakatan antar-*annotator* (*Cohen's Kappa*) tidak dapat diukur. Ketiga, model belum dievaluasi pada data baru di luar periode pengambilan data (*scraping*), sehingga kemampuan generalisasi terhadap ulasan di masa mendatang belum teruji. Keempat, belum terdapat *benchmark* formal untuk *Document Embedding* berbahasa Indonesia, sehingga model *embedding* yang digunakan belum dibandingkan secara sistematis dengan alternatif lain. Keterbatasan tersebut perlu diatasi pada penelitian lanjutan.

Sebagai tindak lanjut strategis dari penelitian ini, manajemen PT KAI dapat mengadopsi model SVM berbasis *Document Embedding* ini sebagai sistem pemantauan sentimen otomatis (*automated monitoring system*) di divisi pelayanan pelanggan atau operasional IT. Langkah ini akan memfasilitasi perusahaan dalam mendeteksi dan menyaring keluhan secara *real time*, sehingga prioritas perbaikan *server backend* dan stabilitas pembayaran dapat dieksekusi lebih cepat berbasis data. Bagi peneliti selanjutnya, kajian analitik ini dapat dikembangkan dengan memperluas jumlah dan sumber data ulasan dari platform lain seperti Apple App Store atau media sosial X (Twitter) guna mendapatkan gambaran opini yang lebih komprehensif. Selain itu, eksplorasi komputasional menggunakan arsitektur *Deep Learning* mutakhir, seperti model sekuensial LSTM (*Long Short-Term Memory*) atau IndoBERT, dapat digunakan untuk menekan angka kesalahan prediksi pada kalimat bersarkasme dan meningkatkan batas maksimal akurasi.

REFERENCES

[1] H. Khalid, E. Shihab, M. Nagappan, and A. E. Hassan, "What do mobile app users complain about?," *IEEE Softw.*, vol. 32, no. 3, pp. 70–77, 2015, doi: 10.1109/MS.2014.50.

[2] R. E. Anderson and S. S. Srinivasan, "E-satisfaction and e-loyalty: A contingency framework," *Psychol. Mark.*, vol. 20, no. 2, pp. 123–138, 2003, doi: 10.1002/mar.10063.

[3] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis sentimen ulasan aplikasi Dana dengan

metode Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 6, no. 9, pp. 4305–4313, 2022.

[4] Q. Le and T. Mikolov, "Distributed Representations of Sentences and Documents," in *Proceedings of the 31st International Conference on Machine Learning*, E. P. Xing and T. Jebara, Eds., in Proceedings of Machine Learning Research, vol. 32. Beijing, China: PMLR, Sep. 2014, pp. 1188–1196. [Online]. Available: https://proceedings.mlr.press/v32/le14.html

[5] D. L. Devi, A. A. Arifiyanti, and S. F. A. Wati, "Analisis Sentimen Ulasan Pengguna Access By Kai Menggunakan Metode Word2vec Dan Algoritma Svm," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4892.

[6] A. Wirayudha, M. Murniyati, and R. Rosdiana, "Analisis Sentimen Terhadap Ulasan Access By KAI Pada Google Play Store Menggunakan Metode Indobert," *Portal Riset dan Inovasi Sistem Perangkat Lunak*, vol. 3, no. 1, pp. 9–20, Jan. 2025, doi: 10.59696/prinsip.v3i1.69.

[7] A. A. Munandar, Farikhin, and C. E. Widodo, "Sentimen analisis aplikasi belajar online menggunakan klasifikasi SVM," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 8, no. 2, 2023, doi: 10.31328/jointecs.v8i2.4747.

[8] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin, and L. Marlinda, "Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," in *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 2018, pp. 1–6. doi: 10.1109/CITSM.2018.8674352.

[9] I. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan E-Commerce Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 8, pp. 302–307, Sep. 2023, doi: 10.30591/jpit.v8i3.5734.

[10] A. Wibowo, F. Noor Hasan, L. Ramadhan, R. Nurhayati, and A. Wibowo, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier," *Jurnal Asiimetrik: Jurnal Ilmiah Rekayasa & Inovasi*, pp. 239–248, Sep. 2022, doi: 10.35814/asiimetrik.v4i1.3577.

[11] N. V Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[12] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, and T. Mikolov, "Learning word vectors for 157 languages," in *Proc. 11th Int. Conf. Lang. Resour. Eval. (LREC)*, Miyazaki, Japan, 2018.

[13] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[14] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[15] A. McCallum and K. Nigam, "A comparison of event models for naive Bayes text classification," in *Proc. AAAI-98 Workshop on Learning for Text Categorization*, 1998, pp. 41–48.

[16] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.

[17] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *Machine Learning: ECML-98*, C. Nédellec and C. Rouveirol, Eds., Berlin, Germany: Springer, 1998, pp. 137–142. doi: 10.1007/BFb0026683.

[18] A. Safira and F. N. Hasan, "Analisis sentimen masyarakat terhadap paylater menggunakan metode Naive Bayes Classifier," *ZONAsi: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, 2023.

[19] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell. (IJCAI)*, Montreal, QC, Canada, 1995, pp. 1137–1145.

[20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: https://doi.org/10.1016/j.patrec.2005.10.010.

[21] Ş. K. Çorbacıoğlu and G. Aksel, "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value," *Turk. J. Emerg. Med.*, vol. 23, no. 4, pp. 195–198, 2023, doi: 10.4103/tjem.tjem_182_23.