

ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP KUALITAS LAYANAN TRANSPORTASI ONLINE PADA APLIKASI KAI ACCESS MENGGUNAKAN METODE DEEP LEARNING DENGAN ALGORITMA LSTM, GRU, DAN CNN

Muhamad Ridwan Usmana, Ahmad Rais Ruli

Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Kwitang, Kec. Senen, Kota Jakarta Pusat 10450, Indonesia
mridwanusmana@gmail.com

ABSTRAK

Perkembangan teknologi informasi dan komunikasi telah mendorong digitalisasi di berbagai sektor, termasuk transportasi online di Indonesia seperti aplikasi KAI Access untuk terus berinovasi dalam memberikan layanan kepada pengguna. Namun, masih banyak pengguna yang menyuarakan ketidakpuasan melalui ulasan di Google Play Store, sehingga diperlukan evaluasi berbasis data terhadap kualitas layanan tersebut. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi KAI Access guna memahami persepsi publik terhadap kualitas layanannya. Metode yang digunakan adalah *Deep Learning* dan pendekatan CRISP-DM dengan algoritma, yaitu *Long Short-Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), dan *Convolutional Neural Network* (CNN). Data dikumpulkan melalui scraping sebanyak 50.000 ulasan dan diproses melalui tahap *pre-pemrosesan* seperti, *case folding*, *cleaning text*, *tokenisasi*, *stopwords*, *stemming*, *normalisasi*, dan pelabelan sentimen. Hasil penelitian menunjukkan bahwa Model GRU memberikan hasil terbaik dengan nilai *accuracy* 98.68% dan AUC paling tinggi 99.01%. Selanjutnya LSTM dan terakhir CNN. Hasil ini menunjukkan bahwa model berbasis *recurrent neural network* lebih efektif dalam memahami urutan kata dan konteks dalam data teks. Selain itu, ditemukan bahwa sentimen negatif masih mendominasi ulasan, terutama terkait fitur pembayaran dan stabilitas aplikasi. Penelitian ini memberikan kontribusi dalam pengembangan model analisis sentimen berbasis Multi-algoritma *deep learning* dan menyajikan rekomendasi bagi pengembang dalam meningkatkan kualitas layanan KAI Access.

Kata kunci : Analisis Sentimen, KAI Access, Deep Learning, LSTM, GRU, CNN, CRISP-DM

1. PENDAHULUAN

Pada era digital teknologi saat ini, perkembangan teknologi informasi dan komunikasi telah membawa perubahan signifikan, sehingga mendorong berbagai sektor untuk bertransformasi secara digital, salah satunya sektor transportasi online di Indonesia. Teknologi informasi dan komunikasi yang berkembang pesat memiliki kemampuan untuk meningkatkan efisiensi dan produktivitas di berbagai sektor[1]. Salah satu sektor transformasi layanan publik dalam transportasi online adalah aplikasi KAI Access. PT Kereta Api Indonesia (Persero) menghadirkan aplikasi KAI Access sebagai inovasi dalam layanan pemesanan tiket dan informasi perjalanan berbasis digital. Aplikasi ini memudahkan pengguna dalam mengakses layanan transportasi khususnya dalam proses pemesanan tiket kereta api secara online [2]. Namun, masih banyak keluhan dari pengguna yang mengindikasikan adanya kekurangan dalam kualitas layanan. Evaluasi terhadap permasalahan ini penting karena akan memberikan pemahaman yang lebih baik tentang peluang dan masalah yang terkait dengan penggunaan teknologi[3].

Dalam konteks ini, pemahaman terhadap opini pengguna menjadi semakin penting, karena dapat memberikan umpan balik langsung terkait pengalaman penggunaan teknologi. Salah satu pendekatan yang dapat digunakan untuk menangkap persepsi tersebut adalah melalui analisis sentimen terhadap ulasan pengguna di platform digital.

Analisis sentimen merupakan pendekatan untuk mengevaluasi kepuasan pengguna terhadap suatu aplikasi, karena dapat menggambarkan pendapat orang secara objektif dari data ulasan teks [4]. Analisis sentimen atau *opinion mining* merupakan bagian dari *natural language processing* (NLP) [5]. Yang berfokus pada identifikasi emosi dan opini pengguna [6].

Seiring dengan meningkatnya volume ulasan pengguna, dibutuhkan model analisis yang mampu memahami konteks bahasa secara mendalam. Proses ini menjadi semakin relevan seiring dengan pesatnya pertumbuhan media sosial, ketersediaan opini dan sentimen publik[7]. Ulasan yang diberikan pengguna merupakan sumber data berharga yang digunakan untuk menggali pola kepuasan atau ketidakpuasan dalam menyusun strategi peningkatan layanan[8].

Penelitian ini dilakukan dengan mempertimbangkan berbagai penelitian sebelumnya yang telah menerapkan metode analisis sentimen dalam konteks ulasan aplikasi. Beberapa penelitian terdahulu telah dilakukan untuk menganalisis ulasan pengguna terhadap aplikasi KAI Access, baik menggunakan pendekatan berbasis *lexicon*, *machine learning*, maupun model berbasis *neural network*. Misalnya, penggunaan model *IndoBERT* [9] menunjukkan akurasi cukup tinggi dalam mengklasifikasikan sentimen pengguna dengan nilai akurasi sebesar 76%, namun belum mengintegrasikan pendekatan multi-algoritma. Studi lainnya

menunjukkan bahwa penggunaan model deep learning yang dapat meningkatkan akurasi klasifikasi sentimen pada data teks yang berbeda dengan metode konvensional, terutama untuk bahasa dengan tingkat kompleksitas tinggi. Meskipun terdapat berbagai pendekatan seperti *Support Vector Machine (SVM)* dan *N-gram* [10], kombinasi analisis sentimen berbasis aspek dengan *naïve Bayes* dan *SVM* [11], dan *Deep Neural Network (DNN)* [12] yang telah digunakan, masih terdapat kesenjangan penelitian (*research gap*) yaitu belum banyak studi yang mengeksplorasi kombinasi algoritma deep learning seperti *Long Short-Term Memory (LSTM)*, *Gated Recurrent Unit (GRU)*, dan *Convolutional Neural Network (CNN)* secara bersamaan. Gap inilah yang menjadi dasar utama untuk mengembangkan pendekatan baru dalam penelitian ini.

Selain kesenjangan dalam integrasi algoritma, masih terdapat ruang untuk pengembangan, terutama pada integrasi multi-algoritma deep learning yang mampu meningkatkan akurasi dan performa klasifikasi sentimen. Selain itu, teknik pra-pemrosesan yang disesuaikan dengan Bahasa Indonesia masih menjadi tantangan tersendiri mengingat kompleksitas struktur bahasa dan keunikan kata dalam bahasa Indonesia.

Dengan integrasi multi-algoritma dan teknik pra-pemrosesan yang disesuaikan dengan bahasa Indonesia, diharapkan penelitian ini dapat memberikan hasil analisis yang lebih akurat dan bermanfaat bagi pengembangan layanan transportasi

online. Keberhasilan penelitian ini memberikan dampak yang signifikan terhadap pemahaman analisis sentimen ulasan pengguna. Penelitian ini juga diharapkan dapat membantu mengembangkan metode analisis sentiment yang dapat digunakan secara luas dalam berbagai konteks aplikasi dengan mengeksplorasi algoritma metode deep learning LSTM, GRU, dan CNN sebagai solusi strategis untuk masalah analisis sentimen terutama ulasan pengguna KAI Access.

Tujuan dari penelitian ini adalah:

- Mengimplementasikan model analisis sentimen yang akurat dari ulasan pengguna KAI Access
- Membandingkan kinerja berbagai algoritma deep learning.
- Menganalisis faktor-faktor yang mempengaruhi sentimen.
- Merumuskan rekomendasi untuk peningkatan layanan.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini dilakukan untuk menjawab berbagai permasalahan yang telah dirumuskan sebelumnya dan mengisi gap yang ditemukan dalam penelitian terdahulu. Penelitian terdahulu berisi studi literatur yang diambil untuk merancang penelitian ini. Adapun penelitian terdahulu terdapat pada Tabel berikut yang merangkum beberapa penelitian terdahulu yang berkaitan dengan analisis sentimen terhadap aplikasi KAI Access.

Tabel 1. Perbandingan Penelitian Terdahulu

No	Peneliti	Metode	Jumlah Data	Hasil Utama	Kelebihan atau keterbatasan
1.	(Wirayudha et al., 2025)	<i>IndoBERT</i>	6.000	Akurasi 76%	Model <i>transformer</i> , namun hanya menggunakan satu algoritma.
2.	(Wardana et al., 2025)	<i>SVM</i> dengan <i>N-Gram</i> dan <i>Glove</i>	1136	<i>SVM</i> dengan fitur <i>Trigram</i> memberikan akurasi tertinggi (84,39%)	Efektif dalam analisis sentimen, <i>GloVe</i> tidak lebih unggul dari <i>N-Gram</i>
3.	(Ariyanti & Susanti, 2024)	<i>Sentiment Analysis + Rating, LSTM</i>	2557	Mayoritas ulasan bintang 1, negatif	Membandingkan performa model menggunakan nilai presisi, <i>recall</i> , akurasi.
4.	(Nugraha & Hendry, 2023)	<i>SVM, Decision Tree, K-NN</i>	1.156	<i>SVM</i> akurasi tertinggi: 88,8%	Tidak menggunakan pendekatan <i>deep learning</i> .
5.	(Sidauruk & Riza, 2023)	Analisis Sentimen literatur, <i>SLR</i>	17 jurnal dalam topik yang sama untuk mengamati tipe perincian model yang diambil.	Sentimen negatif dominan, membandingkan model dari peneliti sebelumnya.	Menggunakan klasifikasi berbasis model <i>machine learning</i> untuk mengetahui model yang sering digunakan.
6.	(Mustakim & Priyanta, 2022)	<i>Aspect-Based + NBC & SVM</i>	3792	Akurasi <i>SVM</i> : 91,63%	Fokus pada aspek.
7.	(Atika et al., 2021)	<i>Deep Neural Network</i>	1000	Klasifikasi sentimen ulasan pengguna	Dataset kecil, belum eksplorasi multi-algoritma.

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa meskipun analisis sentimen terhadap aplikasi KAI Access telah banyak dikaji, pendekatan integritas multi-algoritma deep learning masih jarang digunakan.

2.2. Analisis Sentimen

Analisis sentimen, yang juga dikenal sebagai *mining* opini atau AI emosi merupakan metode Sentimen analisis yang menggunakan pemrosesan bahasa alami (NLP) baik dalam hal analisis teks, identifikasi linguistik, dan mengukur informasi subjektif serta keadaan emosional dari sebuah teks [5]. Subjektivitas adalah kalimat yang menunjukkan perasaan, pandangan, atau keyakinan sedangkan polaritas adalah kalimat yang menunjukkan emosi, yang dapat memiliki nilai positif, negatif, atau netral sedangkan opini adalah kalimat yang menunjukkan perasaan [6].

Analisis sentimen yang merupakan salah satu subbidang dalam pemrosesan bahasa alami, bertujuan untuk mengotomatiskan dan mengkategorikan emosi serta sentimen yang diungkapkan dalam teks tertulis. Proses ini menjadi semakin relevan seiring dengan pertumbuhan pesat media sosial, ketersediaan opini dan sentimen publik semakin melimpah [7].

2.3. Ulasan Pengguna

Ulasan adalah bentuk pendapat atau tanggapan mengenai suatu kejadian, sebagaimana diuraikan dalam Kamus Besar Bahasa Indonesia (KBBI). Secara umum, ulasan dapat dipahami sebagai susunan kalimat atau teks yang berisi evaluasi atau komentar terhadap suatu karya yang dibuat oleh individu [13]. Ulasan tidak hanya berfungsi sebagai sarana untuk menyampaikan opini, tetapi juga dapat menjadi indikator penting dalam menilai kualitas suatu layanan. Dalam konteks digital, ulasan yang diberikan pada platform aplikasi memiliki pengaruh besar terhadap persepsi publik dan keputusan calon pengguna. Oleh karena itu, analisis terhadap ulasan menjadi relevan untuk menggali pemahaman lebih dalam mengenai pengalaman, harapan, dan tingkat kepuasan pengguna terhadap suatu layanan.

2.4. KAI Access

KAI Access adalah aplikasi resmi yang dikembangkan oleh PT Kereta Api Indonesia (Persero) pada tahun 2014 untuk memudahkan seluruh pengguna dalam mengakses layanan transportasi kereta api secara digital [14]. Melalui aplikasi ini, pengguna dapat melakukan pemesanan tiket, pembatalan atau perubahan jadwal perjalanan, check-in mandiri, serta memperoleh informasi *real-time* mengenai jadwal dan status kereta. KAI Access dirancang untuk meningkatkan kenyamanan dan efisiensi pengguna dalam mengelola perjalanan tanpa harus datang langsung ke stasiun, serta mendukung transformasi layanan transportasi publik berbasis teknologi di Indonesia.

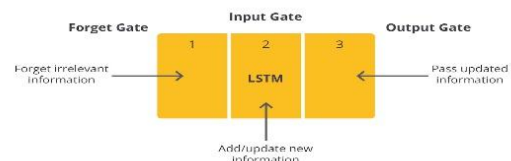
2.5. Deep Learning

Deep learning adalah bagian dari machine learning yang merupakan salah satu teknik machine learning yang canggih serta algoritma yang digunakan adalah jaringan neural buatan yang terinspirasi oleh otak manusia. Keunggulan utama deep learning berbeda dengan metode machine learning yaitu kemampuannya untuk secara otomatis mengekstraksi fitur yang kompleks dari data, dan mampu memprediksi data lebih akurat meskipun harus menggunakan komputasi lebih tinggi dari machine learning [15].

2.6. LSTM

“LSTM, atau *Long Short-Term Memory*, diperkenalkan oleh Sepp Hochreiter dan Jürgen Schmidhuber pada tahun 1997 sebagai solusi untuk mengatasi keterbatasan yang ada pada model jaringan saraf tradisional” [16]. LSTM memiliki struktur yang lebih kompleks dibandingkan dengan *Recurrent Neural Network (RNN)* konvensional, dengan fitur sel memori yang memungkinkan penyimpanan informasi dalam jangka waktu yang panjang serta berbagai mekanisme gating yang mengatur aliran informasi.

2.7. Struktur LSTM



Gambar 1. Struktur LSTM (*Long Short-Term Memory*)

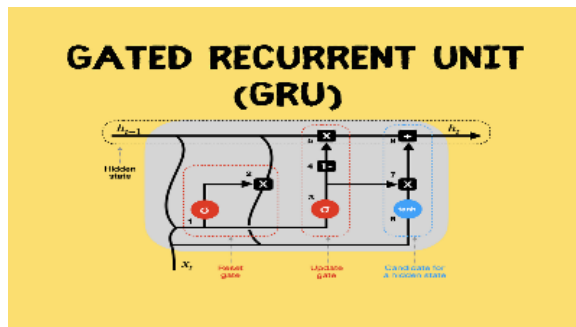
Model ini digunakan untuk menangkap dependensi jangka panjang dalam data teks. Selain itu LSTM digunakan untuk menganalisis urutan kata dalam ulasan dan mengidentifikasi pola sentimen.

2.8. GRU

GRU adalah jenis unit dalam jaringan syaraf rekuren (RNN) yang dirancang untuk memodelkan ketergantungan jangka jauh dalam data urutan dengan cara yang lebih efisien dibandingkan model-model sebelumnya, seperti LSTM. Diperkenalkan oleh Kyunghyun Cho pada tahun 2014, GRU menawarkan alternatif yang lebih sederhana dan lebih hemat dalam penggunaan sumber daya komputasi [17].

2.9. Struktur GRU

GRU (Gated Recurrent Unit) terdiri dari beberapa komponen utama, sebagai berikut:



Gambar 2. Struktur GRU (Gated Recurrent Unit)

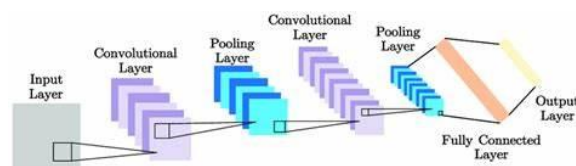
Sementara LSTM memiliki tiga *gate* (*forget gate*, *input gate*, dan *output gate*), GRU hanya memiliki dua *gate* (*update gate* dan *reset gate*), yang membuat struktur GRU lebih sederhana dan lebih efisien secara komputasi. GRU memiliki arsitektur yang lebih sederhana dan dapat lebih cepat dalam pelatihan.

2.10. CNN

Pada tahun 1998, Yann LeCun memperkenalkan *Convolutional Neural Networks* (CNN) untuk pertama kalinya. Dalam bukunya yang berjudul "Pembelajaran Berbasis Gradasi Aplikasi untuk Pengenalan Dokumen", LeCun memperkenalkan versi awal CNN, yang disebut LeNet, yang berhasil mengenali karakter tulisan tangan. Saat ini, LeNet hanya dapat digunakan pada gambar dengan resolusi rendah. Kunihiro Fukushima mengembangkan CNN dengan nama lain NeoCognitron.

CNN biasanya digunakan pada klasifikasi gambar atau dua dimensi, tetapi CNN juga dapat digunakan untuk klasifikasi teks dengan menggunakan konvolusi satu dimensi. Dalam hal ini CNN digunakan untuk menganalisis fitur lokal dalam ulasan dengan menggunakan lapisan konvolusi, model ini akan dapat menangkap pola-pola penting dalam teks yang berhubungan dengan sentimen [18].

Struktur Jaringan Saraf Konvolusional (CNN) terdiri dari beberapa lapisan utama:



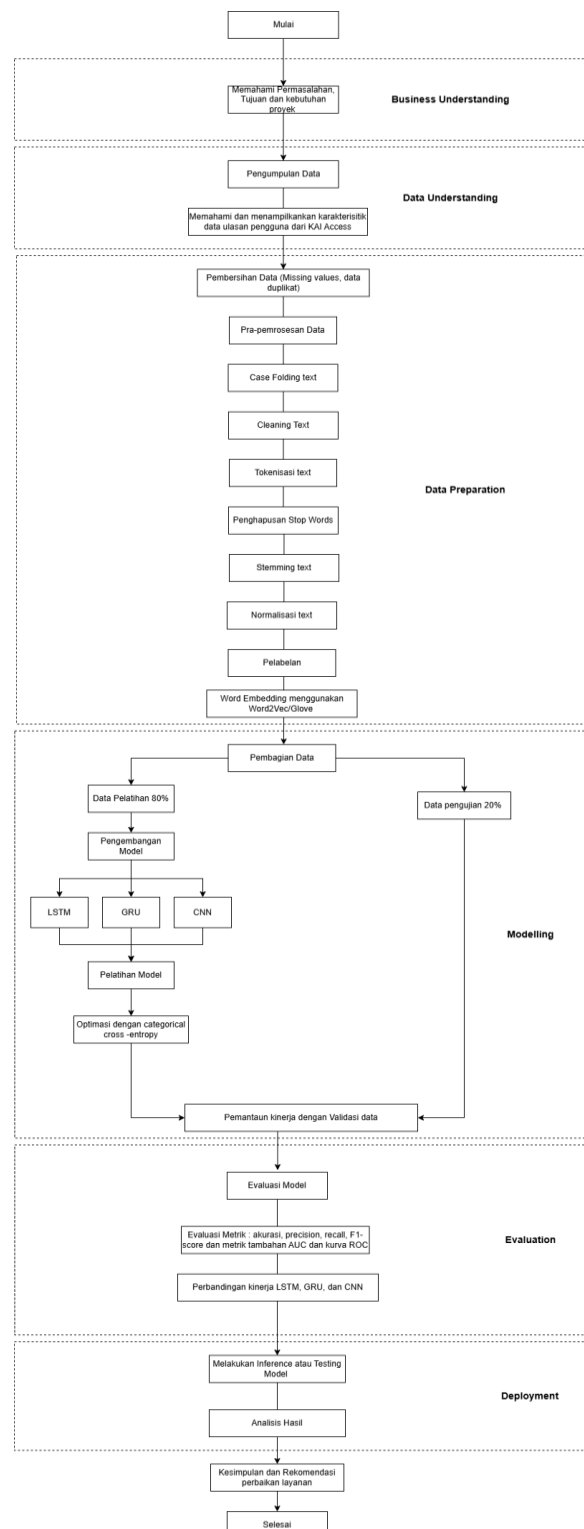
Gambar 3. Struktur CNN (Convolutional Neural Networks)

Dengan demikian, CNN mampu memahami struktur dan konteks bahasa secara bertingkat.

3. METODE PENELITIAN

Tahapan Penelitian merupakan Serangkaian langkah sistematis yang harus dilalui oleh peneliti agar penelitian dapat berlangsung secara terstruktur. Metode penelitian yang digunakan dalam penelitian kuantitatif ini menggunakan klasifikasi sentimen metode deep learning dengan algoritma LSTM, GRU,

dan CNN dan pendekatan yang digunakan dalam penelitian adalah pendekatan eksperimental. Penelitian ini disusun menggunakan kerangka kerja CRISP-DM (*Cross Industry Standard Process for Data Mining*)[19], yang mengorganisir proyek-proyek data science ke dalam enam fase: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [20].



Gambar 4. Tahapan Penelitian

Pada Gambar 4 di atas, terlihat tahapan-tahapan penelitian yang terstruktur yang akan dilakukan, sehingga hal tersebut dapat mempermudah dalam analisisnya.

3.1. Business Understanding

Business Understanding merupakan tahap pemahaman terhadap permasalahan bisnis, penentuan metode pemecahan masalah, serta identifikasi kebutuhan data dan perencanaan proyek manajemen. Pada tahap awal, penelitian berfokus pada pemahaman tujuan dan kebutuhan proyek. Hal ini mencakup identifikasi masalah yang akan diselesaikan, dan penetapan tujuan penelitian secara spesifik.

3.2. Data Understanding

Tahap ini terdiri dari pengumpulan data dan pemahaman karakteristiknya yang bertujuan untuk mendapatkan wawasan di awal:

- Pengumpulan Data, data diperoleh dari ulasan pengguna platform KAI Access. Ulasan yang diambil mencakup berbagai aspek layanan, seperti kemudahan dalam penggunaannya, kecepatan layanan, dan kepuasan pelanggan.
- Pemahaman Data, melakukan eksplorasi data untuk memahami struktur, distribusi, dan karakteristik datasetnya.
- Data yang dikumpulkan dari Google Play Store, yaitu ulasan pengguna dari aplikasi KAI Access yang digunakan untuk penelitian ini. Untuk mendapatkan data ulasan pengguna, dilakukan scraping data menggunakan library `google-play-scraper` yang disediakan oleh pip. Library ini digunakan untuk mengekstrak berbagai fitur dari ulasan aplikasi di Google Play Store. Target pengambilan data adalah 50.000 ulasan data, dan hasilnya disimpan dalam format *Comma Separated Value (.csv)*. Hasil ulasan diurutkan berdasarkan yang terbaru.

3.3. Data Preparation

Data preparation adalah proses pembersihan dan transformasi data mentah menjadi format yang siap digunakan untuk analisis dan pemodelan. Proses ini mencakup pengecekan kualitas data, pembersihan data dari kesalahan seperti data yang hilang, duplikat atau tidak konsisten, serta transformasi data agar sesuai dengan kebutuhan analisis.

Tahap persiapan data meliputi serangkaian proses untuk membersihkan dan mengolah data mentah menjadi format yang siap dianalisis:

- Pembersihan Data yaitu menghilangkan *noise*, karakter khusus, data duplikat, dan data yang *missing values*.
- Pra-pemrosesan Data yaitu mempersiapkan data untuk tahap analisis lanjutan.
- Tokenisasi adalah proses yang memecah teks menjadi bagian yang lebih kecil yang disebut token yang dapat berupa kata, frasa, atau simbol.

d. Case Folding, adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*), sehingga kata seperti "Data" dan "data" dianggap sama dalam analisis. Proses ini membantu mengurangi *kompleksitas* dan *inkonsistensi* dalam data teks.

- Penghapusan Stop Words: proses ini menghilangkan kata-kata umum yang tidak memiliki makna penting secara kontekstual dan tidak memberikan nilai informasi, seperti "dan", "atau", "yang".
- Stemming adalah teknik mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan, awalan, atau akhiran. Contohnya, kata "berlari", "lari-lari", dan "pelari" disederhanakan menjadi "lari".
- Normalisasi Teks yaitu menyeragamkan format teks untuk analisis yang lebih akurat.
- Pelabelan merupakan proses memberikan label atau kategori pada (positif, negatif, dan netral).
- Word Embedding yaitu Teknik representasi kata dalam bentuk vektor numerik Menggunakan teknik *Word2Vec* atau *GloVe* agar dapat diproses oleh algoritma pembelajaran mesin.

3.4. Modeling

Tahap modeling melibatkan pengembangan dan evaluasi model pembelajaran mesin terdiri atas:

- Pembagian Data, data dibagi menjadi dua bagian yaitu data pelatihan (80%) dan data pengujian (20%).
- Pengembangan Model, tiga jenis arsitektur atau algoritma deep learning dikembangkan dan dibandingkan untuk melihat performanya sebagai berikut yaitu Model LSTM, GRU, dan CNN
- Pelatihan Model, setiap model akan dilatih menggunakan data pelatihan dengan menggunakan teknik optimasi. Proses pelatihan akan dilakukan selama beberapa epoch, dengan pemantauan kinerja model pada data validasi untuk menghindari overfitting.
- Optimasi, model dioptimasi menggunakan categorical cross-entropy dan oversampling sebagai fungsi loss yang digunakan untuk klasifikasi sentimen (positif, negatif, netral).
- Validasi, performa model diukur menggunakan data validasi. Proses ini mencakup tuning parameter seperti epoch, learning rate, dan batch size. Validasi silang dilakukan untuk meningkatkan generalisasi dan mengurangi overfitting.

3.5. Evaluation

Tahap Evaluasi ini dilakukan secara komprehensif terhadap model yang telah dikembangkan yaitu Setelah pelatihan, model dievaluasi menggunakan data pengujian. Evaluasi merupakan proses pengamatan dan pertimbangan yang dilakukan kembali sebelum mencapai kesimpulan akhir.

Kinerja model diukur menggunakan metrik seperti akurasi, precision, recall, dan F1-score dengan menggunakan beberapa metrik evaluasi tambahan seperti, AUC dan ROC-curve, untuk mengukur performa masing-masing model secara menyeluruh. Selain itu, dilakukan perbandingan kinerja antara tiga model yaitu LSTM, GRU, dan CNN, guna menentukan model terbaik berdasarkan hasil metrik evaluasi tersebut. Hasil dari ketiga model (LSTM, GRU, dan CNN) digunakan untuk menentukan model mana yang paling efektif dalam menganalisis sentimen ulasan pengguna.

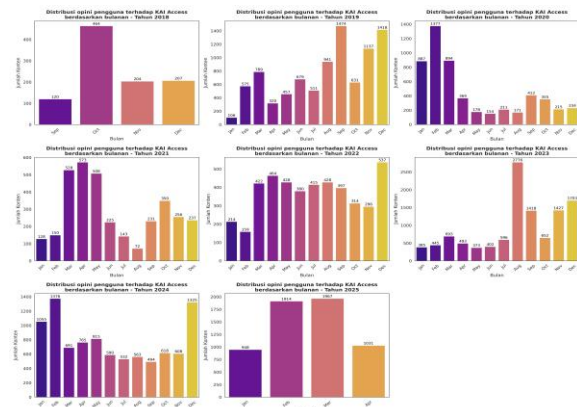
3.6. Deployment

Terakhir yaitu tahap deployment yang mencakup beberapa proses penting, dimulai dengan inferensi atau pengujian model, di mana model terbaik yang telah dipilih digunakan untuk melakukan prediksi pada data baru yang belum pernah dilihat sebelumnya. Selanjutnya, hasil analisis sentimen dianalisis untuk mengidentifikasi tren dan pola dalam ulasan pengguna serta mengetahui sentimen analisisnya berdasarkan hasil visualisasi.

4. HASIL DAN PEMBAHASAN

4.1. Distribusi Ulasan Pengguna bulanan

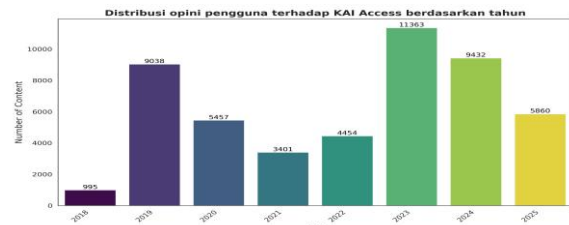
Ulasan pengguna terhadap aplikasi aplikasi KAI Access per bulan untuk setiap tahunnya dari tahun 2018 hingga 2025 menunjukkan pola musiman. Berdasarkan distribusi temporal, pola distribusi bulanan ini menunjukkan variasi yang cukup signifikan setiap tahunnya, yang kemungkinan besar dipengaruhi oleh peristiwa atau musim tertentu seperti libur panjang, momen mudik, atau perubahan layanan. Bisa dilihat pada Gambar 5.



Gambar 5. Distribusi Bulanan Ulasan Pengguna (2018-2025)

Pada Gambar 6, Jumlah ulasan pada awalnya rendah pada tahun 2018 (995 ulasan), tetapi melonjak tajam pada tahun 2019 menjadi 9.038 ulasan, yang mencerminkan peningkatan penggunaan aplikasi. Jumlah ulasan kemudian menurun pada tahun 2020-2022, dengan titik terendah pada tahun 2021 (3.401 ulasan), kemungkinan besar disebabkan oleh dampak pandemi COVID-19. Pada tahun 2023, terjadi

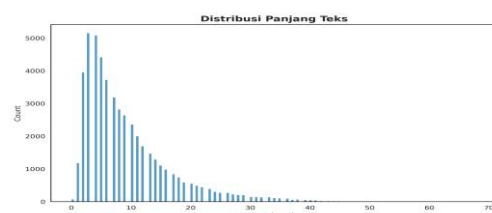
lonjakan dramatis menjadi 11.363 ulasan, yang menandakan pemulihan aktivitas komunitas dan peningkatan partisipasi pengguna. Tahun 2024 menunjukkan sedikit penurunan (9.432 ulasan), dan data untuk tahun 2025 (5.860 ulasan) masih bersifat sementara. Secara keseluruhan, tren ini dipengaruhi oleh faktor sosial, pandemi, dan perkembangan teknologi. Bisa dilihat pada gambar dibawah ini



Gambar 6. Distribusi Tahunan Ulasan Pengguna.

4.2. Karakteristik Teks

Distribusi panjang teks menunjukkan seberapa panjang tulisan atau ulasan yang dibuat oleh pengguna, dan seberapa sering tulisan tersebut muncul. Dari gambar ini, kita dapat melihat bahwa sebagian besar orang menulis ulasan pendek sekitar 2 hingga 10 kata. Semakin panjang ulasannya, semakin sedikit orang yang menulis ulasan tersebut. Sangat jarang ada orang yang menulis lebih dari 30 kata dalam satu ulasan.



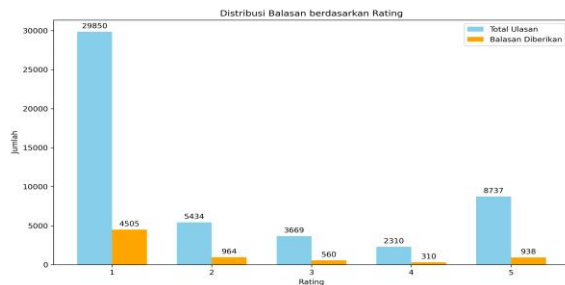
Gambar 7. Distribusi Panjang Teks

Distribusi kata dalam ulasan pengguna KAI Access menunjukkan fokus utama pada penggunaan aplikasi, terutama dalam kaitannya dengan pemesanan dan pembayaran tiket. Kata-kata yang paling sering muncul adalah “aplikasi”, “tiket”, “bayar”, dan “pesan”. Ulasan positif tercermin dalam kata-kata seperti “mudah”, “membantu”, dan “baik”, sementara keluhan pengguna muncul dalam kata-kata seperti “error”, “susah”, dan “lambat”. Secara keseluruhan, kata-kata ini mencerminkan aspek-aspek yang paling dikhawatirkan oleh pengguna dalam pengalaman mereka menggunakan aplikasi. Bisa dilihat pada Gambar 8.



Gambar 8. Kata-kata yang Paling Sering Muncul

Gambar 9. Menunjukkan distribusi balasan KAI access terhadap ulasan pengguna berdasarkan peringkat atau rating (1-5). Tim KAI access lebih sering merespon ulasan negatif (rating 1 dan 2) dengan jumlah sekitar 4505 balasan dan 964 balasan, sedangkan ulasan (rating 3 dan 4) jarang mendapatkan respon, dan ulasan positif (rating 5) yang berjumlah sekitar 8737 juga jarang mendapatkan balasan. Hal ini mengindikasikan bahwa akses KAI lebih fokus pada penanganan keluhan dan umpan balik negatif dari pengguna.

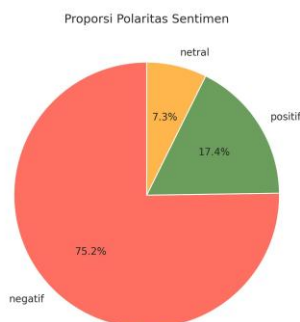


Gambar 9. Distribusi Balasan Berdasarkan Rating

Balasan dari KAI Access sebagian besar terdapat pada rating rendah (rating 1), yang menunjukkan penekanan pada penyelesaian umpan balik negatif.

4.3. Pelabelan

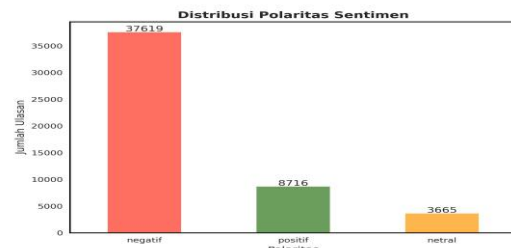
Pada tahap ini, setiap ulasan pengguna diklasifikasikan ke dalam kelas sentimen tertentu, yaitu positif, negatif, atau netral. Proporsi polaritas sentimen ulasan pengguna terhadap layanan KAI Access didominasi oleh sentimen negatif. Berdasarkan diagram lingkaran, 75,2% ulasan bersifat negatif, yang mencerminkan ketidakpuasan pengguna terhadap layanan ini. Sementara itu, 17,4% ulasan bersifat positif dan 7,3% bersifat netral. Temuan ini menunjukkan bahwa persepsi pengguna cenderung negatif, meskipun masih ada ulasan positif terkait aspek-aspek tertentu seperti layanan dan penanganan teknis.



Gambar 10. Proporsi Sentimen Polaritas

Pada Gambar 11, Mayoritas ulasan pengguna KAI Access bernada negatif, yaitu sebanyak 37.619 ulasan, yang mengindikasikan tingkat ketidakpuasan

yang tinggi. Ulasan positif tercatat sebanyak 8.716 ulasan, sedangkan netral sebanyak 3.665 ulasan, yang umumnya berisi saran atau komentar tanpa muatan emosional. Temuan ini mengindikasikan dominasi persepsi negatif terhadap layanan aplikasi.



Gambar 11. Distribusi Polaritas Sentimen

4.4. Visualisasi Wordcloud Sentimen

Untuk memahami kata-kata yang paling sering muncul dalam ulasan pengguna terhadap aplikasi KAI Access, dilakukan visualisasi menggunakan wordcloud pada masing-masing kelas sentimen: positif, negatif, dan netral.

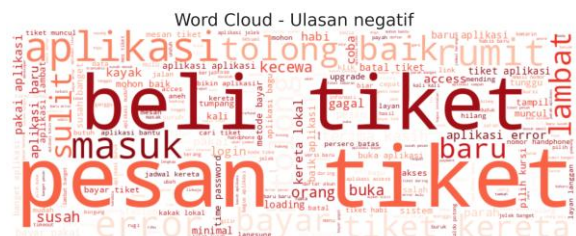
a. Wordcloud Sentimen Positif



Gambar 12. WordCloud Sentimen Positif

Pada wordcloud dengan label sentimen positif, terlihat bahwa kata-kata seperti “mudah”, “cepat”, “praktis”, “bagus”, dan “terima kasih” paling sering digunakan. Ini menunjukkan bahwa pengguna dengan sentimen positif cenderung memberikan apresiasi terhadap kemudahan dan kecepatan layanan dalam aplikasi.

b. Wordcloud Sentimen Negatif



Gambar 13. WordCloud Sentimen Negatif

Sementara itu, pada wordcloud sentimen negatif, kata-kata seperti “error”, “lemot”, “susah”, “gagal”, dan “tidak bisa” mendominasi. Hal ini mengindikasikan bahwa ulasan negatif umumnya berkaitan dengan masalah teknis seperti performa aplikasi yang lambat atau kesulitan dalam melakukan Pemesanan.

c. Wordcloud Sentimen Netral



Gambar 14. WordCloud Sentimen Netral

Untuk sentimen netral, kata-kata seperti “jadwal”, “kereta”, “aplikasi”, dan “info” muncul lebih sering. Kata-kata ini bersifat informatif dan tidak menunjukkan afeksi emosional yang kuat, sesuai dengan karakteristik sentimen netral.

4.5. Performance Model

Berdasarkan seluruh metrik yang digunakan, dilakukan perbandingan performa antara ketiga model. Model GRU secara umum memberikan hasil yang terbaik, dengan akurasi pelatihan, akurasi pengujian, dan nilai AUC yang paling tinggi. LSTM berada di posisi kedua dengan performa yang cukup stabil, sementara CNN tertinggal di belakang, terutama dalam mengidentifikasi sentimen netral. Hasil ini menunjukkan bahwa model berbasis jaringan syaraf tiruan *rekuren* lebih efektif dalam mengidentifikasi sentimen netral. Hasil ini menunjukkan bahwa model berbasis jaringan syaraf tiruan lebih efektif dalam memahami urutan kata dan konteks dalam data teks. Setiap model deep learning (LSTM, GRU, dan CNN) dilatih pada ulasan pengguna yang telah diproses sebelumnya.

Tabel 2. Classification Performance of Each Model

Model	Training Accuracy	Test Accuracy	Accuracy Validation	AUC
LSTM - Word2Vec	0.978900	0.981792	0.980346	0.986344
CNN - Word2Vec	0.964856	0.975368	0.970012	0.981526
GRU - Word2Vec	0.985136	0.986798	0.985887	0.990098

Best Model:

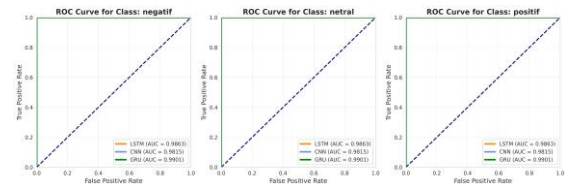
Model: GRU- Word2Vec

Test Accuracy: 0.9868

AUC: 0.9901

Selanjutnya, kurva ROC (*Receiver Operating Characteristic*) dan nilai AUC (*Area Under Curve*) digunakan sebagai evaluasi tambahan yang mengukur kemampuan model untuk membedakan antar kelas. Semakin tinggi nilai AUC, maka semakin baik model tersebut dalam mengklasifikasikan data. Hasil penelitian menunjukkan bahwa model GRU memperoleh nilai AUC tertinggi, diikuti oleh LSTM dan CNN.

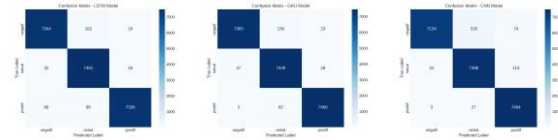
Nilai AUC yang mendekati 1 mengindikasikan bahwa model memiliki sensitivitas dan spesifisitas yang seimbang. Dapat dilihat pada Gambar 15.



Gambar 15. Kurva ROC (LSTM, GRU, dan CNN)

Model GRU mencapai AUC terbaik (0.9901), mengonfirmasi kemampuannya untuk memisahkan kelas sentimen dengan jelas.

Visualisasi Confusion matrix disusun untuk setiap model, yang menunjukkan distribusi prediksi terhadap label aktual. Matriks ini menunjukkan jumlah prediksi yang benar dan salah dari setiap kelas sentimen, sehingga memungkinkan untuk mengidentifikasi jenis kesalahan prediksi yang paling sering terjadi. Sebagai contoh, model CNN menunjukkan tingkat kesalahan tertinggi dalam membedakan antara sentimen netral dan negatif, sementara LSTM cenderung stabil dalam memprediksi sentimen positif.



Gambar 16-18. Confusion Matrix (LSTM, GRU, dan CNN)

LSTM unggul dalam stabilitas klasifikasi. CNN cukup akurat tetapi sensitif terhadap kesalahan antara sentimen negatif dan netral. GRU adalah model dengan kinerja terbaik secara keseluruhan, dengan akurasi tinggi dan kesalahan distribusi yang paling kecil. GRU diklasifikasikan dengan benar: 7.365 negatif, 7.449 netral dan 7.460 ulasan positif dengan kesalahan klasifikasi minimal.

4.6. Prediksi Model dengan Data dari KAI Access
Prediksi Model Dengan User Input

Gambar 19. Prediksi Model Dengan User Input

Tabel 3. Hasil pengujian GRU Model dari KAI Access

	Text	Label Asli	Label Prediksi
0	habis baru offline maksudnya	netral	netral
1	maju	netral	netral
2	tiket habis tampilan listnya bingung satu habi...	positif	positif
3	login sulit login email nomor handphone	netral	netral
4	login	netral	netral
5	login	netral	netral
6	banyak baru	netral	netral
7	baru	netral	netral
8	beli tiket kereta buka lapak hasil bayar pakai...	negatif	negatif
9	aplikasi tanggap bayar kadaluarsa loading masu...	negatif	negatif

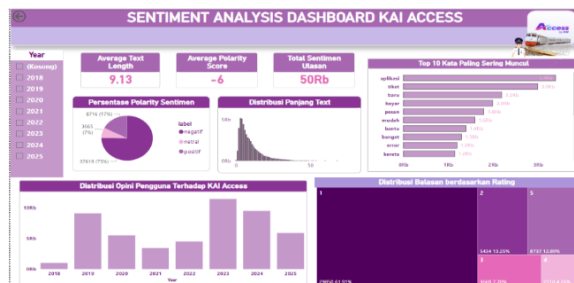
Inference atau testing model dengan dataset aplikasi lain.

Tabel 4. Hasil Testing Model Dengan Dataset E-Commerce

	Kalimat	Prediksi LSTM	Prediksi CNN	Prediksi GRU
0	Produk ini sangat bagus dan berkualitas	positif	positif	positif
1	Saya kecewa dengan pelayanan di sini	negatif	netral	negatif
2	Kualitas biasa saja, tidak terlalu spesial	positif	netral	netral
3	Terima kasih! Sangat senang	positif	positif	positif
4	Pelayanan sangat lambat dan mengecewakan	positif	positif	negatif

4.7. Dashboard Visualisasi Dengan Power BI

Power BI memudahkan dalam menyajikan insight secara *real-time* dan memudahkan stakeholder atau pengambil keputusan dalam memahami tren dan pola persepsi pengguna terhadap layanan KAI Access. Dengan adanya visualisasi ini, proses evaluasi layanan dapat dilakukan dengan lebih cepat, berbasis data, dan efisien.



Gambar 20. Dashboard Sentimen di Power BI

Dashboard ini menampilkan tren sentimen, frekuensi kata, Sentimen berdasarkan tahun, dan memungkinkan pemfilteran berdasarkan kelas dan versi sentimen.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan model analisis sentimen terhadap ulasan pengguna aplikasi KAI Access menggunakan pendekatan deep learning dengan algoritma LSTM, GRU, dan CNN yang mampu mengklasifikasikan sentimen ulasan ke dalam tiga kategori: positif, negatif, dan netral. Dari perbandingan performa ketiga model, GRU terbukti sebagai algoritma terbaik dengan test accuracy mencapai 98,68%, macro F1-score 0,99, dan nilai AUC 0,9901, menunjukkan performa yang lebih superior dibandingkan LSTM yang memperoleh akurasi 98,18% dan CNN dengan akurasi 97,54%.

Penelitian ini juga berhasil mengidentifikasi faktor-faktor yang memengaruhi sentimen pengguna, seperti kemudahan penggunaan, proses pemesanan tiket, kecepatan aplikasi, serta kendala teknis seperti error dan keterlambatan sistem. Melalui analisis distribusi kata dan tren temporal, ditemukan bahwa kata-kata seperti "error", "susah", "lambat" sering muncul pada ulasan negatif, sementara kata "mudah", "bantu", "baik" mendominasi ulasan positif, dengan opini negatif cenderung meningkat pada masa-masa padat seperti musim libur.

Berdasarkan hasil penelitian, disarankan juga agar pihak pengembang aplikasi KAI Access secara proaktif memanfaatkan data ulasan pengguna sebagai dasar perbaikan layanan, terutama dalam menangani keluhan yang berkaitan dengan fitur pembayaran, login, dan stabilitas sistem. Model GRU yang terbukti paling akurat dapat diintegrasikan ke dalam sistem monitoring otomatis untuk mendeteksi opini negatif secara *real-time*, memungkinkan respons yang lebih cepat terhadap keluhan pengguna. Perbaikan layanan yang perlu diprioritaskan meliputi peningkatan performa aplikasi saat periode sibuk, perbaikan pada sistem pembayaran dan pencarian tiket, serta pengembangan dashboard interaktif untuk pengambilan keputusan berbasis data. Framework yang dikembangkan juga berpotensi untuk diterapkan pada sektor layanan publik lainnya yang mengandalkan feedback dari pengguna.

DAFTAR PUSTAKA

- [1] A. Arief and M. Y. Abbas, "Kajian Literatur (Systematic Literature Review): Kendala Penerapan Sistem Pemerintahan Berbasis Elektronik (SPBE)," *PROtek J. Ilm. Tek. Elektro*, vol. 8, no. 1, pp. 1–6, 2021, doi: 10.33387/protk.v8i1.1978.
- [2] K. N. F. Ariyanti and A. Susanti, "Identifikasi

- Pengguna Aplikasi Transportasi Access by KAI dengan Ulasan dan Rating Menggunakan Analisis Sentimen,” *Mitrans J. Media Publ. Terap. Transp.*, vol. 2, no. 1, pp. 31–40, 2024.
- [3] D. Rahmawati, Tjahjanto, and A. R. Ruli, “Literature Review : Implementasi Teknologi Artificial Intelligence dalam Sistem Keamanan Kota di Ibu Kota Nusantara (IKN),” *J. Sist. Inf. dan Apl.*, vol. 2, no. 2, pp. 24–30, 2024.
- [4] N. B. Sidauruk and N. Riza, “SENTIMEN ANALISIS DATA PENGGUNA TERHADAP KAI ACCESS Systematic Literature Review,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 2, pp. 1297–1303, 2023, doi: 10.36040/jati.v7i2.6764.
- [5] D. Jurafsky and J. H. Martin, “Speech and Language Processing,” in *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*, Third Edit., Stanford University & University of Colorado at Boulder, 2025, pp. 1–599. doi: 10.9783/9780812200027.toc.
- [6] Z. Khanam, “Sentiment Analysis of user reviews in an Online Learning Environment: Analyzing the Methods and Future Prospects,” *Eur. J. Educ. Pedagog.*, vol. 4, no. 2, pp. 209–217, 2023, doi: 10.24018/ejedu.2023.4.2.531.
- [7] K. L. Tan, C. P. Lee, and K. M. Lim, “A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research,” *Appl. Sci.*, vol. 13, pp. 1–21, 2023, doi: 10.3390/app13074550.
- [8] D. Hardiansyah, R. A. Aziz, and M. S. Hasibuan, “The Classification Method is Used for Sentiment Analysis in My Telkomsel,” *Int. J. Artif. Intelligence Res.*, vol. 8, no. 2, pp. 1–11, 2024, doi: 10.29099/ijair.v8i2.1229.
- [9] A. Wirayudha, Murniyati, and Rosdiana, “Analisis Sentimen Terhadap Ulasan Access By KAI Pada Google Play Store Menggunakan Metode Indobert,” *PRINSIP Portal Ris. Inov. Sist. Perangkat Lunak*, vol. 3, no. 1, pp. 9–20, 2025.
- [10] R. K. Wardana, N. Cahyono, U. A. Saputro, A. H. Negoro, and N. Aini, “IMPLEMENTASI ALGORITMA SVM DAN OPTIMASI MENGGUNAKAN KOMPARASI N-GRAM DAN GLOVE PADA SENTIMEN PENGGUNA APLIKASI ACCESS BY KAI,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3216–3223, 2025.
- [11] H. Mustakim and S. Priyanta, “Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM,” *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 16, no. 2, pp. 113–124, 2022, doi: 10.22146/ijccs.68903.
- [12] P. D. Atika, Herlawati, F. N. Khasanah, and R. Sari, “Sentiment Analysis of Kai Access Application Using the Deep Neural Network Method,” *IJARCCCE(Int. J. Adv. Res. Comput. Commun. Eng.)*, vol. 10, no. 12, pp. 24–27, 2021, doi: 10.17148/ijarccce.2021.101205.
- [13] F. T. Nugraha and Hendry, “Implementasi Machine Learning Sebagai Analisis Kepuasan Pelanggan Terhadap Penggunaan Aplikasi Kai Access,” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 8, no. 4, pp. 1248–1255, 2023, doi: 10.29100/jupi.v8i4.4185.
- [14] K. KAI Company Profile, “KAI Company Profile. Diambil kembali dari PT Kereta Api Indonesia: https://www.kai.id/static/company-profile/company_profile_2023.pdf,” 2021.
- [15] D. Sanjaya and S. Budi, “Prediksi Pencapaian Target Kerja Menggunakan Metode Deep Learning dan Data Envelopment Analysis,” *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 288–300, 2020, doi: 10.28932/jutisi.v6i2.2678.
- [16] K. Kwanda, D. E. Herwindiati, and M. D. Lauro, “Perbandingan LSTM dan Bidirectional LSTM pada Sistem Prediksi Harga Saham Berbasis Website,” *J. Multidiscip. Res. Dev.*, vol. 7, no. 1, pp. 26–35, 2024.
- [17] Furizal *et al.*, “Long Short-Term Memory vs Gated Recurrent Unit: A Literature Review on the Performance of Deep Learning Methods in Temperature Time Series Forecasting,” *IJRCS Int. Robot. Control Syst.*, vol. 4, no. 3, pp. 1506–1526, 2024.
- [18] J. Zhang *et al.*, “PFLlib : A Beginner-Friendly and Comprehensive Personalized Federated Learning Library and Benchmark,” *J. Mach. Learn. Res.*, vol. 26, pp. 1–10, 2025.
- [19] S. Tripathi, D. Muhr, M. Brunner, H. Jodlbauer, M. Dehmer, and F. Emmert-Streib, “Ensuring the Robustness and Reliability of Data-Driven Knowledge Discovery Models in Production and Manufacturing,” *Front. Artif. Intell.*, vol. 4, pp. 1–20, 2021, doi: 10.3389/frai.2021.576892.
- [20] C. Schröer, F. Kruse, and J. M. Gómez, “A Systematic Literature Review on Applying CRISP-DM Process Model,” *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.