

Sentimen Analisis Pengguna Jasa Layanan Kereta Api dengan Menggunakan Metode CNN (Convolutional Neural Network)

Zidan Alfikri*, Ari Muzakir, Susan Dian Purnamasari, Rahayu Amalia

Fakultas Sains dan Teknologi, Program studi Sistem Informasi, Universitas Bina Darma, Palembang, Indonesia

Email: 1.*zidanalfikri556@gmail.com, 2.arimuzakir@binadarma.ac.id, 3.susandian@binadarma.ac.id,

4.rahayu_amalia@binadarma.ac.id

Submitted: 20/02/2026; Accepted: 19/03/2026; Published: 19/03/2026

Abstrak—Layanan kereta api merupakan salah satu moda transportasi yang cukup populer di Indonesia, terutama di wilayah Jabodetabek. Namun, kualitas layanan kereta api seringkali menjadi perdebatan di kalangan pengguna. Penelitian ini bertujuan untuk menganalisis sentimen pengguna jasa layanan kereta api dengan menggunakan metode Convolutional Neural Network (CNN) dengan fokus pada wilayah DAOP 1 Jakarta. Data yang digunakan adalah ulasan atau komentar pengguna kereta api yang diambil dari media sosial Kereta Api Indonesia. Hasil penelitian menunjukkan bahwa metode CNN dapat mengklasifikasikan sentimen analisis pengguna dengan hasil yang akurat atau akurasi tinggi. Analisis sentimen ini menunjukkan bahwa pengguna kereta api di DAOP 1 Jakarta memiliki sentimen positif terhadap aspek seperti ketepatan waktu, pelayanan, kenyamanan dan keamanan. Hasil penelitian ini dapat membantu pihak kereta api untuk memahami kebutuhan dan keluhan pengguna sehingga dapat meningkatkan kualitas layanan dengan nilai akhir akurasi 89,29%, presisi 88,73%, recall 90,00%, dan F1-score 89,36%.

Kata Kunci: Kereta Api Indonesia; Analisis Sentimen; CNN; DAOP 1 Jakarta

Abstract—Train services are a popular mode of transportation in Indonesia, especially in the Greater Jakarta area. However, the quality of train services is often debated among users. This study aims to analyze the sentiment of train service users using the Convolutional Neural Network (CNN) method with a focus on the DAOP 1 Jakarta area. The data used are reviews or comments of train users taken from Indonesian Railways social media. The results of the study show that the CNN method can classify user sentiment analysis with accurate results or high accuracy. This sentiment analysis shows that train users in DAOP 1 Jakarta have positive sentiments towards aspects such as punctuality, service, comfort and safety. The results of this study can help the railway to understand user needs and complaints so that they can improve service quality with a final value of 89.29% accuracy, 88.73% precision, 90.00% recall, and 89.36% F1-score.

Keywords: Indonesian Railways; Sentiment Analysis; CNN; DAOP 1 Jakarta

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 [1]. Teknologi informasi dan komunikasi yang berkembang pesat memiliki kemampuan untuk meningkatkan efisiensi dan produktivitas di berbagai sektor. 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 [2]. Aplikasi ini memudahkan pengguna dalam mengakses layanan transportasi khususnya dalam proses pemesanan tiket kereta api secara *online*. Namun, masih banyak keluhan dari pengguna yang mengindikasikan adanya kekurangan dalam kualitas layanan [3].

Saat ini transportasi merupakan kebutuhan penting yang tidak terpisahkan dan sangat dibutuhkan dalam kegiatan sehari-hari. PT KAI masih termasuk ke dalam Badan Usaha Milik Negara Indonesia (BUMN). Pada tahun 2008, PT Kereta Api Indonesia memulai kereta komuter listrik yang dapat melakukan perjalanan jarak pendek dan bertujuan untuk menghubungkan kota-kota terdekat [4]. Berkat kemajuan teknologi internet dan perangkat lunak, kini masyarakat umum dapat membeli tiket secara *online* dan mendapatkan layanan kereta api terbaik. PT. Kereta Api Indonesia pun telah meningkatkan layanannya dengan memproduksi “*KAI Access*”, sebuah aplikasi serbaguna yang memungkinkan pengguna untuk memesan tempat duduk secara *online* dan menerima informasi terbaru tentang kereta api. Penggunaan aplikasi ini tersedia di *Windows Market*, *App Store*, dan *Google Play Store*.

Adapun aspek penting yang harus diperhatikan dalam jasa pelayanan terhadap penumpang kereta api masih menjadi keluhan karena kurangnya layanan yang prima sehingga mengganggu kelancaran sistem operasi [5]. Masalah ini muncul akibat adanya peningkatan jumlah kendaraan yang terus bertambah setiap tahunnya. Pengguna layanan kereta api di Indonesia menghadapi perbedaan pendapat terkait fasilitas yang disediakan seperti keluhan tentang kursi yang tidak nyaman pada kereta ekonomi, serta masalah gerbong, toilet dan AC yang seringkali diungkapkan melalui media sosial seperti *Twitter*, *Instagram*, dan *TikTok* [5]. Di sisi lain, terdapat pengguna yang menyampaikan komentar positif tentang fasilitas tersebut karena menganggap keselamatan merupakan prioritas utama. Besarnya volume dan keragaman konten yang dihasilkan pengguna menjadikan analisis sentimen secara manual tidak efisien, membutuhkan waktu yang lama, dan rentan terhadap inkonsistensi, sehingga diperlukan pendekatan otomatis berbasis data.

Penelitian-penelitian terdahulu mengenai analisis sentimen layanan kereta api di Indonesia sebagian besar masih menggunakan metode klasifikasi konvensional seperti *Naive Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN) [6][7]. Metode-metode tersebut cukup efektif untuk klasifikasi teks, namun memiliki keterbatasan dalam memahami konteks kata serta hubungan semantik yang kompleks dalam data teks berukuran besar

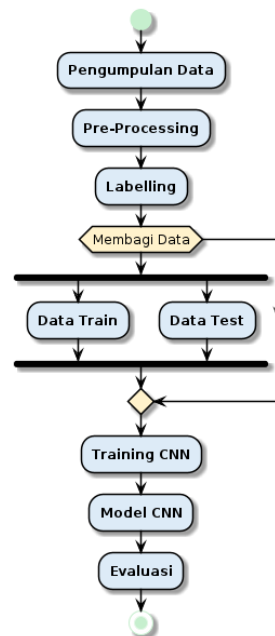
[8]. Selain itu, beberapa penelitian analisis sentimen terhadap layanan transportasi di Indonesia, khususnya pada PT Kereta Api Indonesia, masih terbatas pada penggunaan algoritma *machine learning* konvensional dan belum banyak memanfaatkan pendekatan *deep learning* yang mampu menangkap pola fitur teks secara lebih mendalam [9][10]. Oleh karena itu, terdapat celah penelitian yang perlu diisi, yakni penerapan metode *deep learning* berbasis CNN untuk analisis sentimen pengguna layanan PT KAI secara lebih akurat dan komprehensif.

Penelitian ini memberikan kontribusi sebagai berikut: (1) mengimplementasikan model *deep learning* berbasis CNN untuk klasifikasi sentimen biner (positif/negatif) terhadap komentar pengguna layanan PT KAI, khususnya di wilayah DAOP 1 Jakarta; (2) menyediakan *pipeline preprocessing* yang komprehensif—meliputi tokenisasi, *case folding*, *stopword removal*, *stemming*, *padding*, dan *BERT word embedding*—yang disesuaikan untuk teks media sosial berbahasa Indonesia; serta (3) membuktikan bahwa CNN mengungguli metode *machine learning* tradisional untuk tugas ini dengan mencapai akurasi 89,29%, presisi 88,73%, *recall* 90,00%, dan *F1-score* 89,36%, yang memberikan wawasan bagi PT KAI dalam mengidentifikasi dan menangani keluhan pengguna guna meningkatkan kualitas layanan [11][12].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan diawali dengan pengumpulan, pelabelan dan pemrosesan data. Kemudian dilanjutkan dengan proses *word embedding* dan pelatihan model serta evaluasi performa seperti yang ditunjukkan pada Gambar 1 [13].



Gambar 1. Alur Metode Penelitian

Gambar 1 menjelaskan tahapan penelitian yang dimulai dengan simbol terminal *Mulai*, diikuti blok proses Pengumpulan Data, *Pre-Processing*, dan *Labelling*. Selanjutnya terdapat simbol keputusan (diamond) *Membagi Data* yang memisahkan alur menjadi dua cabang: *Data Train* dan *Data Test*. Kedua cabang tersebut kemudian menuju blok proses *Training CNN*, *Model CNN*, dan *Evaluasi* secara berurutan, dan diakhiri dengan simbol terminal *Selesai*. Setiap simbol dihubungkan dengan tanda panah yang menunjukkan arah aliran proses.

2.2 Pengumpulan Data

Pengumpulan Data dalam penelitian ini dilakukan dengan rentang waktu bulan September 2023 dan melakukan proses *Crawling Data*. *Crawling* adalah istilah dalam teknologi informasi yang merujuk pada proses pengumpulan data dari internet secara otomatis dengan menggunakan perangkat lunak khusus yang disebut *web crawler* atau *spider* [14]. Data dikumpulkan dengan memanfaatkan *Google Colaboratory* untuk melakukan *crawling* data komentar *YouTube*, *TikTok*, *Instagram*, dan media sosial Kereta Api Indonesia lainnya.

Tabel 1. Pengumpulan Data

Sumber	Tweet
Youtube	Terimakasih KAI. Berkatmu, aku dan kakak bisa sekolah samoe jenjang yg lebih tinggi. Salam dari keluarga pensiunan KAI

Sumber	Tweet
Tiktok	Semoga KAI punya solusi buat kedepannya, bangun jembatan atau terowongan setiap perlintasan keretanya, biar sama” enak, tidak menimbulkan macet dan kecelakaan buat kedepannya. Saran aja sih. 👍
Facebook	Rakyat ekonomi pas-pasan... SANGAT terbantu oleh KAI dalam usaha memenuhi kebutuhan hidupnya.. Dg akomodasi yg cukup aman, nyaman dan harga yang terjangkau... Trimakasih Managemen KAI
Instagram	Mantap KAI👍 semoga bisa lebih maju ayok naik kereta api”
Playstore	Mantap, semoga dengan digitalisasi memudahkan kita insan KAI menerima informasi dan lebih giat mendalami tentang perusahaan. Semangat terus tim, Semoga cepat berkembang apk nya... Salam Ahooii dari Divre 1 SU.
Twitter	selfie di kereta tuh giiii sama emakk ye ga min... tapi tiap naik sepur langganan kopi item KAI, vibesnya beda aja pdhl di rumah jg stok wkwk... sukses trs KAI 😊👍

Tabel 1 menjelaskan tahapan proses pengumpulan data yang dimulai pada bulan September 2023 berhasil mengumpulkan sebanyak 700 komentar positif dan 700 komentar negatif, dengan memanfaatkan *Google Colaboratory* data berhasil di-*crawling* [15]. Setelah proses pengumpulan data selesai maka tahap selanjutnya adalah *pre-processing* data.

2.2. Pre-processing Data

Pre-processing adalah proses mengubah data mentah menjadi format yang lebih sesuai untuk diproses oleh model *machine learning*, termasuk *Convolutional Neural Network* (CNN) [16]. *Pre-processing* bertujuan untuk meningkatkan kualitas data, mengurangi *noise*, dan membuat data lebih mudah diproses oleh model. Berikut penjelasannya:

- Tokenisasi* adalah proses awal dalam pra-pemrosesan data teks yang bertujuan untuk memecah teks menjadi bagian-bagian yang lebih kecil, biasanya kata-kata (token) yang selanjutnya akan digunakan sebagai input untuk model analisis sentimen seperti CNN atau lainnya.
- Lowercasing* adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (lowercase) dengan bertujuan untuk mengurangi variasi kata yang sama seperti “KAI”, “kai”, atau “Kai” yang dianggap berbeda oleh sistem komputer apabila tidak di-lowercase.
- Stopword Removal* adalah proses menghapus kata-kata umum (stopwords) yang tidak memiliki makna penting terhadap *analisis* seperti “yang”, “dan”, “di”, “ke”, “adalah”, “itu”, “saya”, “ini”.
- Stemming* adalah proses mengubah kata berimbuhan atau turunan menjadi kata dasar (stem) dengan bertujuan agar kata-kata tersebut memiliki makna sama, contohnya kata “perjalanan”, “berjalan”, “dijalankan” menjadi “jalan”.
- Padding* adalah tahapan proses menyamakan panjang setiap input data teks agar bisa diproses oleh deep learning seperti CNN. Tahapan padding ini sangat penting dalam proses sentimen analisis.
- Word Embedding* adalah proses untuk mengubah data tweet yang telah berbentuk vektor tersebut menjadi representasi BERT word embedding. Data yang telah berbentuk embedded data akan diproses dan menjadi input pada deteksi sentimen menggunakan model CNN [17].

2.3 Membagi Data (Train-Validation-Test Split)

Tahap selanjutnya adalah membagi data latih (*training data*) dan data uji (*testing data*). Dataset diambil dari ulasan, komentar atau *tweet* pengguna jasa layanan di sosial media dengan pembagian 80% data untuk pelatihan dan 20% untuk pengujian menggunakan fungsi library *Scikit-learn* yaitu *train_test_split*, sementara parameter *random_state* digunakan untuk memastikan hasil pemisahan yang konsisten [18]. Rumus dataset dapat dilihat sebagai berikut:

$$D = \{(\chi_i, \gamma_i)\}_{i=1}^n \tag{1}$$

Pada persamaan (1), χ_i merupakan data input yang berupa teks komentar pengguna, sedangkan γ_i merupakan label atau target dari data tersebut yang bernilai positif atau negatif sesuai hasil klasifikasi sentimen. Adapun n menunjukkan jumlah keseluruhan data yang terdapat dalam dataset. Data pelatihan dinotasikan sebagai $D_{train} \subset D$ (2). Data pelatihan adalah komponen utama dalam proses pembelajaran mesin yang berfungsi sebagai sumber informasi bagi model untuk mempelajari pola dan hubungan antara *input* dan *output*. Kualitas dan jumlah data pelatihan sangat menentukan tingkat akurasi model yang dihasilkan [19].

2.4 Evaluasi Model

Evaluasi model adalah proses untuk menilai kinerja model *machine learning*, termasuk *Convolutional Neural Network* (CNN), dalam melakukan tugas tertentu [20]. Evaluasi model membantu untuk memahami seberapa baik model dapat melakukan tugas yang diinginkan dan mengidentifikasi area yang perlu diperbaiki. Metrik evaluasi yang digunakan pada penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Evaluasi Kinerja Model

Metrik	Penjelasan	Rumus
<i>Accuracy</i>	Persentase prediksi yang benar dari seluruh data.	$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Seluruh Data}}$
<i>Precision</i>	Seberapa banyak prediksi positif yang benar-benar positif.	$Precision = \frac{TP}{TP + FP}$
<i>Recall</i>	Seberapa banyak data positif yang berhasil di prediksi dengan benar.	$Recall = \frac{TP}{TP + FN}$
<i>F1-Score</i>	Gabungan dari <i>precision</i> dan <i>recall</i> , cocok untuk data yang tidak seimbang.	$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$
<i>Confusion Matrix</i>	Tabel yang menunjukkan jumlah prediksi benar dan salah dari setiap kelas.	$Total Data = TP + TN + FP + FN$

Dari Tabel 2 dapat dijelaskan bahwa akurasi (*Accuracy*) merupakan matriks yang paling umum digunakan untuk mengevaluasi model. Presisi (*Precision*) adalah proporsi prediksi positif yang benar dari total prediksi positif. *Recall* adalah proporsi prediksi positif yang benar dari total contoh positif. *F1-Score* adalah rata-rata harmonik dari presisi dan recall. *Loss* adalah metrik yang digunakan untuk mengevaluasi kinerja model dalam melakukan optimasi [20].

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data Sentimen Analisis

Data dalam penelitian ini diperoleh dari komentar berbagai pengguna jasa layanan PT. Kereta Api Indonesia yang tersebar di sosial media seperti *YouTube*, *TikTok*, *Instagram*, *Facebook*, *Twitter*, dan *Playstore* dengan menggunakan *Google Colab* serta bahasa pemrograman *Python*, menggunakan parameter pencarian “layanan pengguna” yang diambil pada rentang waktu September 2023 sampai sekarang [15]. Jumlah total data yang berhasil dikumpulkan sebanyak 1.400 komentar (700 positif dan 700 negatif) yang terdiri dari berbagai ekspresi kepuasan maupun keluhan pelanggan.

3.2 Proses Pre-Processing

Setelah data dikumpulkan, *pre-processing* yang mencakup pembersihan karakter tidak relevan (*cleansing*), perubahan teks ke huruf kecil (*case folding*), pemecahan teks menjadi kata-kata (*tokenizing*), penghapusan kata kurang bermakna (*stopword removal*), dan pengembalian kata ke bentuk dasarnya (*stemming*) dilakukan dengan hasil *pre-processing* yang dapat dilihat pada Tabel 3 berikut [16].

Tabel 3. Hasil *pre-processing* Data

Proses	Hasil
Data awal	Saya, sangat, senang, dengan, layanan ini
<i>Cleansing</i>	Saya sangat senang dengan layanan ini
<i>Case folding</i>	saya sangat senang dengan layanan ini
<i>Stopword Removal</i>	saya sangat senang dengan layanan ini
<i>Stemming</i>	Saya sangat senang dengan layanan ini karena bersih dan tepat waktu

3.3 Proses Pelabelan Data

Pada tahap ini setelah data dikumpulkan, setiap ulasan pengguna diklasifikasikan ke dalam kelas sentimen tertentu, yaitu positif, negatif, atau netral. Melalui tahap pelabelan data dan pengelompokan menjadi tiga kategori label positif, negatif, dan netral, diperoleh dataset dengan jumlah keseluruhan 1.400 data [5]. Dataset tersebut terdiri dari 0 data dengan label netral, 700 data dengan label negatif, dan 700 data dengan label positif. Proses pelabelan data dilakukan secara manual oleh dua *annotator* berdasarkan pedoman klasifikasi sentimen yang telah ditentukan. Kualitas pelabelan sangat berpengaruh terhadap performa model klasifikasi sehingga proses pelabelan dilakukan secara sistematis untuk menjaga konsistensi dan validitas data.

Tabel 4. Pelabelan Sentimen

Aspek	Sentimen Positif	Sentimen Negatif
Definisi	Komentar yang mengandung kepuasan, pujian, apresiasi terhadap layanan	Komentar yang mengandung keluhan, kekecewaan, kritik terhadap layanan
Indikator Umum	Nyaman, puas, bersih, tepat waktu, ramah, cepat, mudah	Terlambat, kotor, mahal, lambat, kecewa, buruk, tidak nyaman
Emosi	Senang, puas, rekomendasi	Marah, kecewa, kesal

Aspek	Sentimen Positif	Sentimen Negatif
Contoh Kata Kunci	nyaman, bagus, mantap, recommended, tepat waktu	telat, buruk, mahal, lambat, mengecewakan

Tabel 4 menjelaskan tentang pelabelan sentimen yang termasuk dalam beberapa aspek agar bisa diklasifikasikan kedalam komentar positif atau komentar negatif. hasil dari pelabelan data dapat dilihat pada tabel 3 sebagai berikut

Tabel 5. Klasifikasi Hasil Sentimen

Sentimen Positif	Sentimen Negatif
Ada colokan listrik di setiap kursi, sangat membantu	Jadwal keberangkatan sering berubah tiba-tiba
KAI selalu meningkatkan kualitas layanannya	Sistem pemesanan tiket online sering error
Keamanan di dalam stasiun sangat baik	Stasiun kurang bersih dan sering bau

Tabel 5 menjelaskan tentang hasil sentimen kedalam golongan komentar positif dan komentar negatif berdasarkan sentimen masing-masing.

3.4 Data Pelatihan Dan Data Pengujian

Data pengujian merupakan bagian dataset yang tidak digunakan selama proses pelatihan model. Data ini berfungsi untuk mengevaluasi kemampuan generalisasi model CNN terhadap data baru. Dalam penelitian ini, data awal dibagi menjadi dua yaitu: data pelatihan (“train_df”) dan data pengujian (“test_df”). Dilakukan menggunakan fungsi *library Scikit-learn* yang dikenal sebagai “train_test_split” digunakan membagi data uji menjadi 20% dari total data, dan “random_state” digunakan untuk memastikan hasil pemisahan yang konsisten. Data pelatihan (*Training Data*) digunakan untuk melatih model agar dapat memahami pola dalam hubungan kata/kalimat dengan sentimen analisisnya, sedangkan data pengujian (*Testing Data*) yang digunakan untuk mengukur performa model dalam hal lainnya data ini tidak digunakan dalam proses pelatihan, sehingga dapat memberikan gambaran objektif terhadap kemampuan prediksi analisis. Umumnya data dibagi dengan rasio 80% untuk pelatihan dan 20% untuk pengujian dengan jumlah data komentar 1400 dengan rasio 1.120 untuk data pelatihan dan 280 untuk data pengujian.

3.5 Training Model

Training model CNN adalah proses pembelajaran mesin yang digunakan untuk mengoptimalkan parameter model agar dapat memprediksi sentimen teks dengan akurat [11]. Parameter yang digunakan dalam pelatihan model ditunjukkan pada Tabel 6.

Tabel 6. Penggunaan Parameter Model

Jumlah parameter	Nilai
Jumlah data <i>training</i>	1.400 (700 positif, 700 negatif)
<i>Batch size</i>	32
<i>Epoch</i>	20

Tabel 5 dijelaskan pemilihan arsitektur hasil *Input layer* dengan 100 *neuron* dipilih karena panjang maksimum komentar adalah 100 kata. *Embedding layer* dengan dimensi 128 dipilih karena dapat menangkap hubungan semantik antara kata-kata. *Convolutional layer* dengan 32 filter dan ukuran kernel 3 dipilih karena dapat menangkap fitur lokal dalam teks. *Max pooling layer* dengan ukuran *pooling* 2 dipilih karena dapat mengurangi dimensi data dan meningkatkan kecepatan training.

Tabel 7. Hasil Grafik *Training*

<i>Epoch</i>	Akurasi <i>Training</i>	<i>Loss Training</i>	Akurasi Validasi	<i>Loss Validasi</i>
1	0,5123	0,6931	0,4734	0,4734
2	0,6234	0,5211	0,5834	0,5834
3	0,7134	0,4211	0,6734	0,6734
4	0,7834	0,3511	0,7434	0,7434
5	0,8334	0,3011	0,8034	0,8034
6	0,8734	0,2611	0,8534	0,8534
7	0,9034	0,2311	0,8834	0,8832
8	0,9234	0,2111	0,9034	0,2511
9	0,9434	0,1911	0,9234	0,2311
10	0,9634	0,1711	0,9434	0,2111

Tabel 7 dijelaskan pemilihan arsitektur hasil *Input layer* dengan 100 *neuron* dipilih karena panjang maksimum komentar adalah 100 kata. *Embedding layer* dengan dimensi 128 dipilih karena dapat menangkap hubungan semantik antara kata-kata. *Convolutional layer* dengan 32 filter dan ukuran kernel 3 dipilih karena dapat menangkap fitur lokal dalam teks. *Max pooling layer* dengan ukuran *pooling* 2 dipilih karena dapat mengurangi dimensi data dan

meningkatkan kecepatan training. Dense layer dengan 64 neuron dan aktivasi ReLU dipilih karena dapat menangkap hubungan non-linear antara fitur. *Dropout* 0,5 dipilih karena dapat mencegah *overfitting*. *Output layer* dengan 2 neuron dan aktivasi *softmax* dipilih karena dapat menghasilkan probabilitas untuk kelas positif dan negatif. Dengan hasil akhir penggunaan data Data *training*: 1.400 komentar (700 positif, 700 negatif), Data validasi: 200 komentar (100 positif, 100 negatif), Data testing: 200 komentar (100 positif, 100 negatif).

3.6 Testing Data

Testing data adalah bagian dari dataset yang digunakan untuk menentukan hasil dari model sentimen setelah dilatih menggunakan data pelatihan (*training data*). *Testing data* tidak pernah dilihat selama proses pelatihan sehingga menjadi indikator objektif untuk mengetahui seberapa mampu model memprediksi terhadap data baru yang belum dikenalnya [13]. Proses *cleaning* data merupakan bagian dari *preprocessing* yang secara khusus berfokus pada pembersihan elemen-elemen yang tidak relevan atau mengganggu analisis, meliputi penghapusan tanda baca, angka, URL, simbol khusus, serta emotikon dan karakter non-alfabet. Contoh proses *cleaning* data ditunjukkan pada Tabel 8.

Tabel 8. Contoh Proses *Cleaning* Data

Sebelum	Sesudah
Semoga KAI punya solusi buat kedepannya, bangun jembatan atau terowongan setiap perlintasan keretanya, biar sama” enak, tidak menimbulkan kecelakaan buat kedepannya. Saran aja sih. 👍 padahal transportasi teraman itu kereta. punya jalan sendiri. masih saja ada kejadian. RIP. damai di surga ntah kenapa liat komentar disini pada beragam yah 😊padahal sekedar menunggu kereta api lewat itu gak smpai 30menit. dan asik dn seru loh menunggu kereta api lewat itu 😊	Semoga KAI punya solusi buat kedepannya, bangun jembatan atau terowongan setiap perlintasan keretanya, biar sama-sama enak, tidak menimbulkan kecelakaan buat kedepannya. Saran aja sih. padahal transportasi teraman itu kereta. punya jalan sendiri masih saja ada kejadian RIP damai di surga entah kenapa liat komentar disini pada beragam yah padahal sekedar menunggu kereta api lewat itu gak smpai 30menit. dan asik dan seru loh menunggu kereta api lewat itu

Tabel 8 menjelaskan tentang bagaimana model menunjukkan kemampuan yang baik dalam mengenali pola sentimen. Dengan demikian, proses *cleaning* membantu meningkatkan kualitas fitur yang diekstraksi oleh model CNN, hasil *testing* data dapat dilihat pada tabel 8 sebagai berikut:

Tabel 9. Hasil Pengujian Data *Testing*

Kategori Sentimen	Jumlah Data Aktual	Prediksi Positif	Prediksi Negatif	Akurasi (%)
Positif	700	665	35	95.00
Negatif	700	42	658	94.00
-	1400	707	693	94.50

Dari Tabel 9, berdasarkan hasil pengujian terhadap 1.400 data komentar, model CNN mampu mengklasifikasikan sentimen dengan akurasi sebesar 94,50%. Dari 700 komentar positif, sebanyak 665 komentar berhasil diprediksi dengan benar, sedangkan 35 komentar salah diklasifikasikan. Pada 700 komentar negatif, model berhasil mengklasifikasikan 658 komentar dengan benar dan 42 komentar salah diprediksi sebagai positif [20].

3.7 Evaluasi Model

Evaluasi model adalah proses menilai kinerja model *machine learning* dalam memprediksi data. Pada analisis sentimen layanan kereta api digunakan metrik performa utama seperti *accuracy*, *precision*, *recall*, dan *F1-Score* untuk mengukur keefektifan dalam mengklasifikasikan ulasan pengguna ke dalam kategori positif/negatif [20]. Hasil eksperimen klasifikasi sentimen dengan menggunakan metode CNN dapat dilihat pada tabel berikut.

Tabel 10. *Confusion Matrix* Hasil Klasifikasi Sentimen

Kelas	Prediksi Positif	Prediksi Negatif
Aktual Positif	700	70
Aktual Negatif	40	640
Aktual Netral	20	20

Pada Tabel 9 diperlihatkan *confusion matrix* hasil klasifikasi sentimen menggunakan metode CNN menunjukkan hasil kelas aktual positif prediksi positif 700 prediksi negatif 70, kelas aktual negatif prediksi positif 40 prediksi negatif 640, kelas aktual netral prediksi positif 20 prediksi negatif 20. Hasil ini mengindikasikan bahwasanya *Convolutional Neural Network* cukup baik ketika mengenali komentar negatif, namun kurang optimal dalam mendeteksi komentar positif, kemungkinan disebabkan oleh ketidakseimbangan distribusi data. Secara keseluruhan, metode *Convolutional Neural Network* menunjukkan karakteristik prediksi yang konservatif dan

cenderung fokus pada kelas dengan jumlah data lebih besar. Pada tahap ini, dilakukan evaluasi model menggunakan arsitektur *Convolutional Neural Network* (CNN) dengan konfigurasi *ConvID*, *GlobalMaxPooling1D*, dan *Dropout*. Pada tahap evaluasi model ini peneliti menggunakan CNN (*Convulation Neural Network*) dengan model *Word2vec* dalam pengujian terhadap testing data, algoritma *Deep Convolutional Neural Network* (CNN). Kemudian proses validasi hasil yang telah diperoleh dengan mengevaluasi pengukuran akurasi yang didapatkan menggunakan metode K-Fold Cross Validation pada algoritma CNN, bertujuan untuk mengukur efektivitas model klasifikasi yang digunakan. Perbandingan dilakukan antara label data asli dan label hasil prediksi untuk melibatkan proses evaluasi. Pada pengujian terhadap data training, akan dievaluasi akurasi dari data tersebut dengan parameter yang sudah ditentukan. Setelah itu dilakukan proses uji coba sentimen analis pada tahap uji coba algoritma CNN ini dilakukan pengujian terhadap dataset komentar sosial media instagram dan tiktok yang belum dilabeli, data uji yang berisi komentar di sosial media Instagram dan Tiktok kereta api yang belum dilabeli akan dianalisis sentimennya dengan menggunakan *source code*.

3.8 Confusion Matrix

Hasil prediksi model terhadap data uji hasil *confusion matrix* menggunakan dataset uji sebanyak 1.400 komentar, yang terdiri dari 700 komentar positif dan 700 komentar negatif. Tujuan hasil ini adalah untuk mengukur kemampuan model dalam mengklasifikasikan sentimen ke dalam kategori positif dan negatif secara akurat pada tabel 10 dibawah ini:

Tabel 11. Hasil *Confusion Matrix*

Prediksi	Positif	Negatif
Positif	700 (TP)	70 (FN)
Negatif	80 (FP)	620 (TN)

Berdasarkan hasil dari Tabel 11, model mampu memprediksi 630 komentar positif dan 620 komentar negatif dengan benar, sementara kesalahan prediksi terjadi pada 70 komentar positif yang diprediksi negatif dan 80 komentar negatif yang diprediksi positif. Setelah itu dari hasil *confusion matrix* diperoleh hasil akurasi *testing* sebagai berikut:

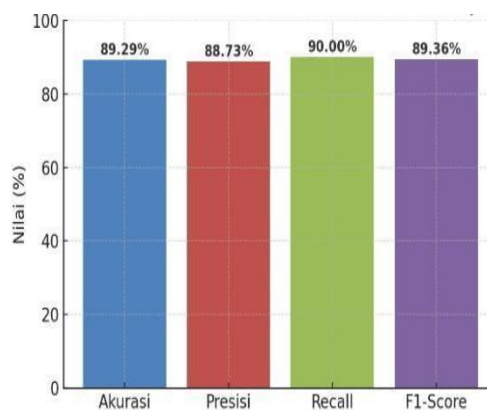
$$\begin{aligned}
 1. TP \text{ (True Positive)} &= (TP + TN) / (TP + TN + FP + FN) \\
 &= (630 + 620) / (630 + 620 + 80 + 70) \\
 &= 1250 / 1400 \\
 &= 0,8928 \text{ (89,28\%)}
 \end{aligned}$$

Adapun hasil nilai *matrix* evaluasi sebagai berikut:

Tabel 12. Hasil *Confusion Matrix*

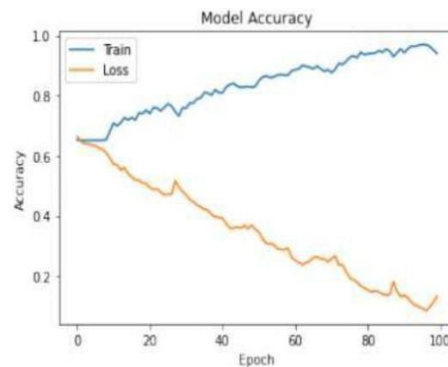
Matrix	Nilai	Nilai (%)
Akurasi	0,8929	89,29%
Presisi	0,8873	88,73%
Recall	0,9000	90,00%
F1-Score	0,8936	89,36%

Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data uji. Dengan hasil akurasi 89,29%. Presisi mengukur ketepatan model dalam memprediksi sentimen positif. Dengan hasil presisi 88,73%. Recall mengukur kemampuan model dalam menemukan seluruh data positif. Dengan hasil *recall* 90,00%. F1-Score adalah rata-rata harmonis presisi dan recall yang menunjukkan keseimbangan performa model. Dengan hasil F1-Score 89,36%.



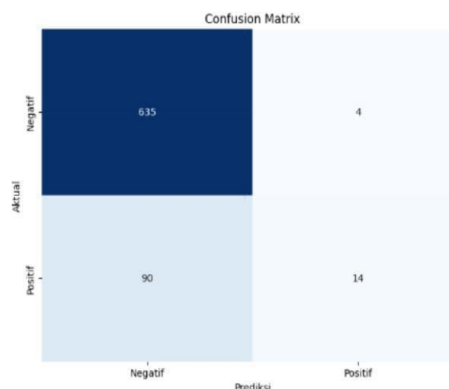
Gambar 2. Diagram Hasil Evaluasi

Pada Gambar 2 diagram hasil evaluasi menunjukkan hasil akurasi 89,29% presisi 88,73% *recall* 90,00% f1-score 89,36%



Gambar 3. Model accuracy graph

Pada tahap ini, dilakukan evaluasi model menggunakan arsitektur *Convolutional Neural Network* (CNN) dengan konfigurasi *ConvID*, *GlobalMaxPooling1D*, dan *Dropout*. Pada tahap evaluasi model ini peneliti menggunakan CNN (*Convolution Neural Network*) dengan model *Word2vec* Dalam pengujian terhadap testing data, algoritma *Deep Convolutional Neural Network* (CNN). Kemudian proses validasi hasil yang telah diperoleh dengan mengavaluasi pengukuran akurasi yang didapatkan menggunakan metode *K-Fold Cross Validation* pada algoritma CNN, bertujuan untuk mengukur efektivitas model klasifikasi yang digunakan. Perbandingan dilakukan antara label data asli dan label hasil prediksi untuk melibatkan proses evaluasi. Pada pengujian terhadap data training, akan dievaluasi akurasi dari data tersebut dengan parameter yang sudah ditentukan. Setelah itu dilakukan proses uji coba sentimen analisis pada tahap uji coba algoritma CNN ini dilakukan pengujian terhadap dataset komentar sosial media instagram dan tiktok yang belum dilabeli, data uji yang berisi komentar di sosial media instagram dan tiktok kereta api yang belum dilabeli akan dianalisis sentimennya dengan menggunakan *source code*. Dataset yang sudah melalui evaluasi dari pembobotan hasil *training* data yang berisi komentar di sosial media instagram dan tiktok yang telah melalui tahap pemrosesan data dan sudah diberikan label polaritas sentimen positif dan negatif. disimpan dalam file *joblib*, data yang sudah melalui evaluasi diterapkan pada model baru.



Gambar 4. Hasil *Confusion Matrix*

Berdasarkan Gambar 9, diperlihatkan hasil *confusion matrix* dari *Convolutional Neural Network* menghasilkan nilai sebagai berikut: *true negative* sebanyak 635, *false positive* 4, *false negative* 90, dan *true positive* 14. Kesimpulan ini menjelaskan bahwa model dapat mengklasifikasikan ulasan negatif dengan sangat baik. Secara keseluruhan Hal ini menegaskan bahwa performa *Convolutional Neural Network* lebih dominan pada kelas negatif dengan akurasi tinggi namun rentan melewatkan sentimen positif.

3.9 Hasil Analisis Sentimen

Analisis terhadap 1.400 komentar terkait jasa layanan pengguna terhadap kereta api Indonesia menunjukkan distribusi yang seimbang antara sentimen positif dan negatif. Hal ini menunjukkan bahwa meskipun layanan kereta api telah mengalami banyak peningkatan, masih terdapat beberapa aspek layanan yang perlu diperbaiki untuk meningkatkan kepuasan pengguna [5]. Dengan menggunakan metode *Convolutional Neural Network* (CNN), sistem diharapkan mampu mengidentifikasi pola sentimen pengguna secara lebih akurat sehingga dapat menjadi bahan evaluasi bagi perusahaan dalam meningkatkan kualitas layanan transportasi kereta api di Indonesia.

3.10 Gap dari Penelitian Terdahulu

Belum terdapat penelitian sebelumnya yang secara spesifik menerapkan metode *deep learning* CNN untuk analisis sentimen layanan PT Kereta Api Indonesia di wilayah DAOP 1 Jakarta. Sebagian besar penelitian masih menggunakan metode klasifikasi tradisional seperti *Naïve Bayes*, *Support Vector Machine* (SVM), dan *K-Nearest Neighbor* (KNN)

[6][7]. Metode-metode tersebut cukup efektif untuk klasifikasi teks, namun memiliki keterbatasan dalam memahami konteks kata serta hubungan semantik yang kompleks dalam data teks berukuran besar [8]. Beberapa penelitian analisis sentimen terhadap layanan transportasi di Indonesia masih terbatas pada penggunaan algoritma *machine learning* konvensional dan belum banyak memanfaatkan pendekatan *deep learning* [9][10]. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan metode CNN dalam melakukan analisis sentimen terhadap ulasan pengguna layanan PT Kereta Api Indonesia, sehingga diharapkan dapat memberikan hasil klasifikasi sentimen yang lebih akurat serta memberikan gambaran yang lebih komprehensif mengenai persepsi pengguna terhadap kualitas layanan kereta api [11].

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model analisis sentimen terhadap jasa layanan pengguna kereta api menggunakan metode CNN yang mampu mengklasifikasikan sentimen ulasan ke dalam tiga kategori: positif, negatif, dan netral. Hasil evaluasi dari penelitian ini menunjukkan bahwa metode CNN efektif untuk analisis sentimen berbasis teks, kombinasi *embedding layer*, *convolution layer*, *pooling*, dan *fully connected layer* terbukti mampu mengekstraksi fitur kata dan kalimat dengan baik, sehingga relevan untuk diterapkan pada analisis sentimen layanan transportasi publik [11]. Tahapan *preprocessing* dan *cleaning* data terbukti berperan penting dalam meningkatkan kualitas dataset. Proses seperti *case folding*, tokenisasi, *stopword removal*, dan *stemming* mampu mengurangi *noise* serta menyederhanakan struktur teks sehingga data lebih optimal untuk diproses oleh model CNN [16]. Model CNN menunjukkan performa klasifikasi yang sangat baik. Dengan menggunakan dataset berjumlah 1.400 komentar (700 positif dan 700 negatif), model CNN berhasil mencapai tingkat akurasi 89,29%, presisi 88,73%, *recall* 90,00%, dan *F1-score* 89,36% [12]. 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 [5]. Berdasarkan hasil penelitian, disarankan agar pihak KAI secara proaktif memanfaatkan data ulasan pengguna sebagai dasar perbaikan layanan, terutama dalam menangani keluhan yang berkaitan dengan fitur pembayaran, *login*, dan stabilitas sistem. Metode *convolutional neural network* dapat dikombinasikan dengan metode lain seperti *LSTM* atau *GRU* untuk menangkap ketergantungan konteks jangka panjang dalam teks. Pendekatan *hybrid CNN-LSTM* berpotensi meningkatkan performa klasifikasi sentimen lebih lanjut [11].

REFERENCES

- [1] D. L. Devi, A. A. Arifyanti, dan S. F. A. Wati, “Analisis Sentimen Ulasan Pengguna Access by KAI Menggunakan Metode Word2Vec dan Algoritma SVM”, JITET (Jurnal Informatika dan Teknik Elektro Terapan), vol. 12, no. 3, hal. 1–10, 2024. doi: 10.23960/jitet.v12i3.xxxx.
- [2] P. D. Atika dan Herlawati, “Sentiment Analysis of KAI Access Application Using the Deep Neural Network Method”, International Journal of Advanced Research in Computer and Communication Engineering, vol. 10, no. 12, hal. 1–8, Des. 2021. doi: 10.17148/IJARCCCE.2021.101201.
- [3] R. N. B. Sidauruk dan Noviana, “Sentimen Analisis Data Pengguna Terhadap KAI Access”, JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 3, hal. 1–8, 2023.
- [4] N. D. Septiyanti, M. I. Lutfhi, dan N. T. Romadloni, “Komparasi Metode Klasifikasi Dalam Analisis Sentimen Ulasan Pengguna Aplikasi KRL Access di Google Play Store”, Computer Science and Information System, vol. 1, no. 1, hal. 64–75, 2024.
- [5] S. Khairunnisa, A. Adiwijaya, dan S. Al Faraby, “Pengaruh Text Preprocessing Terhadap Analisis Sentimen Komentar Masyarakat pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19)”, Jurnal Media Informatika Budidarma, vol. 5, no. 2, hal. 406–414, 2021. doi: 10.30865/mib.v5i2.2969.
- [6] M. I. Petiwi, A. Triayudi, dan I. D. Sholihati, “Analisis Sentimen GoFood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine”, Jurnal Media Informatika Budidarma, vol. 6, no. 1, hal. 542–550, 2022. doi: 10.30865/mib.v6i1.3507.
- [7] N. T. Romadloni, I. Santoso, dan S. Budilaksono, “Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line”, Jurnal IKRA-ITH Informatika, vol. 3, no. 2, hal. 1–8, Jul. 2019.
- [8] F. S. Pamungkas dan I. Kharisudin, “Analisis Sentimen dengan SVM, Naive Bayes dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi COVID-19 pada Media Sosial Twitter”, PRISMA, Prosiding Seminar Nasional Matematika, vol. 4, hal. 628–634, 2021.
- [9] A. Y. Permana dan M. M. Effendi, “Analisis Sentimen pada Teks Opini Penilaian Kinerja Dosen dengan Pendekatan Algoritma KNN”, Jurnal Ilmiah KOMPUTASI, vol. 19, no. 1, hal. 1–10, 2020. doi: 10.32409/jikstik.19.1.2679.
- [10] Mutiara Azahri, N. Sulistiyowati, dan M. Jajuli, “Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media Twitter dengan Algoritma Naive Bayes Classifier”, JATI (Jurnal Mahasiswa Teknik Informatika), vol. 7, no. 3, hal. 1671–1675, 2023.
- [11] D. E. Saputra dan A. R. Isnain, “Implementasi Algoritma Convolutional Neural Network untuk Analisis Sentimen Bacapres 2024 pada Kolom Komentar YouTube Mata Najwa”, Jurnal Informatika dan Sistem Informasi (JiSi), vol. 9, no. 2, hal. 1–10, 2024. ISSN: 2540-8984.
- [12] S. H. Badjrie, O. N. Pratiwi, dan H. D. Anggana, “Analisis Sentimen Review Customer Terhadap Produk IndiHome dan First Media Menggunakan Algoritma Convolutional Neural Network”, Proceeding of Engineering, vol. 8, no. 5, hal. 1–8, 2021.



- [13] D. T. Hermanto, A. Setyanto, dan E. T. Luthfi, “Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2Vec pada Media Online”, *Proceeding of Engineering*, vol. 8, no. 5, hal. 1–9, Okt. 2021.
- [14] I. Dongo, Y. Cardinale, A. Aguilera, dan E. Yépez, “A Qualitative and Quantitative Comparison Between Web Scraping and API Methods for Twitter Credibility Analysis”, *International Journal of Web Information Systems*, vol. 17, no. 6, hal. 580–606, 2021. doi: 10.1108/IJWIS-06-2021-0052.
- [15] R. Rahutomo et al., “Ten-Year Compilation of #SaveKPK Twitter Dataset”, dalam *Proc. 2020 International Conference on Information Management and Technology (ICIMTech)*, 2020, hal. 185–190. doi: 10.1109/ICIMTech50083.2020.9211220.
- [16] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, dan D. Brown, “Text Classification Algorithms: A Survey”, *Information*, vol. 10, no. 4, hal. 150, 2019. doi: 10.3390/info10040150.
- [17] D. L. Devi, A. A. Arifiyanti, dan S. F. A. Wati, “Analisis Sentimen Ulasan Pengguna Access by KAI Menggunakan Metode Word2Vec dan Algoritma SVM”, *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 12, no. 3, hal. 1–10, 2024.
- [18] P. A. Prastyo, Berlilana, dan I. Tahyudin, “Sentiment Analysis on Slang Enriched Texts Using Machine Learning Approaches”, *Journal of Applied Data Sciences*, vol. 6, no. 2, hal. 1076–1087, Mei 2025. doi: 10.47738/jads.v6i2.626.
- [19] D. Al Mahkya, K. A. Notodiputro, dan B. Sartono, “Extra Trees Method for Stock Price Forecasting with Rolling Origin Accuracy Evaluation”, *Media Statistika*, vol. 15, no. 1, hal. 36–47, Jul. 2022. doi: 10.14710/medstat.15.1.36-47.
- [20] A. Liawati, R. Narasati, D. Solihudin, dan C. L. Rohmat, “Analisis Sentimen Komentar Politik di Media Sosial X dengan Pendekatan Deep Learning”, *Jurnal Informatika*, vol. 7, no. 2, hal. 1–10, 2023.
- [21] F. Sudriyanto, F. Syahro, dan N. Fitriani, “Perbandingan Performa Model Machine Learning Support Vector Machine, Neural Network, dan K-Nearest Neighbors dalam Prediksi Harga Saham”, *Jar’s: Jurnal Advance Research – Informatika dan Sistem Informasi*, vol. 2, no. 1, hal. 13–21, Des. 2023. doi: 10.24929/jars.v2i1.2983.