Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

Analisis Sentimen Ulasan Pengguna Halodoc Menggunakan TextCNN pada Dataset Tidak Seimbang Berbasis NLP

Fitriyani[*], Budi Tjahjono

Ilmu Komputer, Teknik Informatika, Universitas Esa Unggul, Jakarta, Indonesia
Email: [1,*]fi3djanie@student.esaunggul.ac.id, [2]budi.tjahjono@esaunggul.ac.id
Email Penulis Korespondensi: fi3djanie@student.esaunggul.ac.id
Submitted: 01/02/2026; Accepted: 22/08/2026; Published: 31/08/2026

Abstrak–Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Halodoc menggunakan arsitektur TextCNN pada dataset berbahasa Indonesia yang memiliki ketidakseimbangan kelas ekstrem. Dataset terdiri dari 2.000 ulasan yang dikumpulkan dari Google Play Store dan dilabeli secara manual ke dalam tiga kelas, yaitu positif, negatif, dan netral. Distribusi data menunjukkan 1.800 ulasan positif (90%), 175 negatif (8,75%), dan 25 netral (1,25%). Untuk mengurangi bias terhadap kelas mayoritas, model TextCNN dilatih menggunakan teknik class weighting dan dievaluasi menggunakan accuracy, precision, recall, F1-score, serta confusion matrix. Hasil pengujian memperoleh accuracy sebesar 93,75% dan weighted F1-score sebesar 0,94, tetapi macro F1-score hanya sebesar 0,57. Recall kelas netral mencapai 0%, menunjukkan bahwa model belum mampu mengenali kelas minoritas secara efektif. Kontribusi penelitian ini mencakup tiga aspek: evaluasi empiris TextCNN pada ulasan Halodoc berbahasa Indonesia dengan ketidakseimbangan kelas ekstrem; analisis dampak distribusi kelas melalui metrik per kelas dan error analysis; serta identifikasi implikasi praktis hasil klasifikasi bagi pemantauan kualitas layanan kesehatan digital. Temuan ini menunjukkan bahwa accuracy yang tinggi tidak cukup untuk merepresentasikan kinerja model pada dataset tidak seimbang, sehingga strategi seperti oversampling, data augmentation, atau pendekatan hibrida perlu dipertimbangkan pada penelitian selanjutnya.

Kata Kunci: Analisis Sentimen; TextCNN; Natural Language Processing; Layanan Kesehatan Digital; Halodoc

Abstract–This study analyzes user sentiment in Halodoc application reviews using a TextCNN architecture on an Indonesian-language dataset with extreme class imbalance. The dataset consists of 2,000 reviews collected from the Google Play Store and manually labeled into three classes: positive, negative, and neutral. The class distribution comprises 1,800 positive reviews (90%), 175 negative reviews (8.75%), and 25 neutral reviews (1.25%). To reduce bias toward the majority class, the TextCNN model was trained using class weighting and evaluated using accuracy, precision, recall, F1-score, and a confusion matrix. The model achieved an accuracy of 93.75% and a weighted F1-score of 0.94, while the macro F1-score was only 0.57. The recall for the neutral class was 0%, indicating that the model was unable to recognize the minority class effectively. The contributions of this study are threefold: an empirical evaluation of TextCNN on Indonesian Halodoc reviews with extreme class imbalance; an analysis of the effect of class distribution using class-level metrics and error analysis; and an identification of the practical implications of the classification results for monitoring digital healthcare service quality. These findings demonstrate that high accuracy alone is insufficient to represent model performance on imbalanced datasets; therefore, strategies such as oversampling, data augmentation, or hybrid approaches should be considered in future research.

Keywords: Sentiment Analysis; TextCNN; Natural Language Processing; Digital Healthcare Services; Halodoc

1. PENDAHULUAN

Analisis sentimen terhadap ulasan pengguna merupakan salah satu pendekatan Natural Language Processing (NLP) yang dapat digunakan untuk memahami pengalaman, persepsi, dan tingkat kepuasan pengguna terhadap layanan kesehatan digital. Ulasan pada platform seperti Halodoc mengandung informasi mengenai pelayanan dokter, aplikasi, obat, konsultasi, pembayaran, serta pengalaman pengguna lainnya. Jumlah ulasan yang terus bertambah membuat analisis manual menjadi kurang efisien, sehingga diperlukan metode otomatis yang dapat mengelompokkan opini ke dalam sentimen positif, negatif, dan netral. Perkembangan NLP di bidang kesehatan menunjukkan bahwa pengolahan teks dapat mendukung analisis umpan balik pasien dan pengambilan keputusan berbasis data [1], [2], [3]. Namun, karakteristik ulasan pengguna yang pendek, tidak terstruktur, dan menggunakan bahasa sehari-hari membuat proses klasifikasi sentimen tetap menjadi tantangan.

Salah satu tantangan penting dalam penelitian ini adalah ketidakseimbangan distribusi kelas. Dataset yang digunakan terdiri dari 2.000 ulasan, dengan 1.800 ulasan positif, 175 negatif, dan hanya 25 netral. Dengan demikian, kelas positif mencakup 90% data, sedangkan kelas netral hanya 1,25%. Kondisi tersebut dapat menyebabkan model lebih mudah mempelajari pola kelas mayoritas dan mengabaikan pola kelas minoritas. Akibatnya, accuracy dapat terlihat tinggi meskipun model gagal mengenali kelas yang jumlahnya sedikit. Permasalahan ini relevan dalam analisis data kesehatan karena ulasan negatif dan netral justru dapat mengandung informasi penting mengenai aspek layanan yang perlu diperbaiki. Oleh sebab itu, evaluasi perlu mempertimbangkan precision, recall, F1-score, dan terutama macro F1-score, bukan hanya accuracy.

Penelitian terdahulu menunjukkan bahwa berbagai pendekatan telah digunakan untuk analisis sentimen pada domain kesehatan. Sidabutar dan Juardi (2025) menganalisis 5.000 ulasan Google Play Store mengenai Halodoc menggunakan Support Vector Machine (SVM) untuk klasifikasi positif dan negatif. Penelitian tersebut menunjukkan bahwa SVM dapat menghasilkan akurasi tinggi, tetapi fokusnya hanya pada dua kelas dan pendekatan berbasis fitur tradisional memiliki keterbatasan dalam menangkap konteks kalimat yang kompleks

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

[4]. Imaduddin, A'la, dan Nugroho (2023) menggunakan IndoBERT pada 9.310 ulasan beberapa aplikasi kesehatan, yaitu Halodoc, Alodokter, dan KlikDokter, dan melaporkan accuracy 96%, precision 95%, recall 96%, serta F1-score 95% [5]. Penelitian tersebut menunjukkan keunggulan representasi kontekstual Transformer, tetapi objeknya mencakup beberapa aplikasi dan hanya menggunakan kelas positif dan negatif.

Penelitian lain memperlihatkan perkembangan model deep learning untuk teks kesehatan. Al-Hadhrami et al. (2024) mengevaluasi Bi-LSTM dan Bi-LSTM-CNN pada ulasan obat pasien dan melaporkan bahwa model Bi-LSTM-CNN dapat mencapai accuracy 96% pada salah satu skenario pengujian [6]. Temuan tersebut menunjukkan bahwa arsitektur hibrida mampu menangkap informasi sekuensial dan pola lokal, tetapi konteks datanya berupa ulasan obat, bukan ulasan aplikasi telemedicine Indonesia. Ouni, Benmohamed, dan Ltifi (2024) mengusulkan SoftEMB yang menggabungkan representasi GloVe dan Word2Vec melalui soft voting pada beberapa arsitektur deep learning berbasis CNN dan RNN. Studi tersebut menunjukkan peningkatan accuracy pada beberapa dataset sentimen [7]. Meskipun demikian, penelitian tersebut tidak secara khusus mengevaluasi dampak ketidakseimbangan kelas ekstrem pada tiga kelas sentimen ulasan layanan kesehatan.

Studi terbaru juga menunjukkan bahwa NLP dan machine learning semakin banyak digunakan untuk mengolah umpan balik layanan kesehatan. Li et al. (2025) mengembangkan sistem untuk mengklasifikasikan keluhan pasien menggunakan NLP dan beberapa algoritma machine learning. Dataset penelitian tersebut tidak seimbang dan diseimbangkan menggunakan SMOTE; SVM mencapai accuracy rata-rata 0,91 pada data eksternal [3]. Hasil tersebut menegaskan bahwa penanganan ketidakseimbangan data merupakan bagian penting dalam klasifikasi teks kesehatan. Ye, Chang, dan Fan (2025) mengembangkan NE ZHA-TextCNN dengan penambahan representasi konteks global untuk klasifikasi teks panjang dan memperoleh performa tinggi pada dataset medis [8]. Temuan ini menunjukkan bahwa TextCNN dapat diperkuat dengan informasi konteks, terutama ketika teks memiliki struktur dan informasi yang lebih kompleks.

Berdasarkan penelitian-penelitian tersebut, terdapat beberapa research gap yang menjadi dasar penelitian ini. Pertama, penelitian Halodoc oleh Sidabutar dan Juardi menggunakan SVM dan dua kelas sentimen, sedangkan penelitian ini menggunakan TextCNN untuk tiga kelas sehingga kelas netral dapat dianalisis secara eksplisit [4]. Kedua, penelitian Imaduddin et al. menunjukkan performa tinggi IndoBERT pada beberapa aplikasi kesehatan, tetapi tidak berfokus pada satu aplikasi dan tidak menekankan persoalan ketidakseimbangan ekstrem antara kelas positif, negatif, dan netral [9]. Ketiga, penelitian Al-Hadhrami et al. dan Ouni et al. membuktikan efektivitas deep learning serta representasi embedding, tetapi menggunakan domain dan distribusi data yang berbeda [7], [6]. Keempat, penelitian Li et al. (2025) menunjukkan pentingnya strategi penyeimbangan data pada klasifikasi keluhan kesehatan, sedangkan penelitian ini secara khusus menguji batas kemampuan class weighting pada dataset dengan kelas netral hanya 1,25% [3]. Kelima, penelitian Ye et al. (2025) memperlihatkan bahwa pengayaan konteks dapat meningkatkan TextCNN, tetapi belum membahas secara khusus ulasan aplikasi kesehatan berbahasa Indonesia yang sangat pendek dan tidak seimbang [8]. Dengan demikian, masih terdapat kebutuhan untuk mengevaluasi seberapa jauh TextCNN dengan class weighting mampu mempertahankan kinerja pada kelas minoritas ketika diterapkan pada ulasan Halodoc dengan ketidakseimbangan ekstrem.

Penelitian ini menggunakan TextCNN karena arsitektur tersebut mampu mengekstraksi pola lokal berbasis n-gram dengan struktur yang relatif ringan dan sesuai untuk klasifikasi teks. Dataset penelitian terdiri dari 2.000 ulasan Halodoc berbahasa Indonesia yang dikumpulkan dari Google Play Store pada 5–6 Juli 2025 dan dibagi menjadi tiga kelas sentimen. Teknik class weighting digunakan untuk memberikan penalti lebih besar terhadap kesalahan pada kelas minoritas. Evaluasi tidak hanya melihat accuracy, tetapi juga precision, recall, F1-score, macro F1-score, weighted F1-score, dan confusion matrix untuk memperlihatkan performa setiap kelas.

Kontribusi penelitian ini terdiri atas tiga aspek. Pertama, penelitian memberikan evaluasi empiris TextCNN pada ulasan Halodoc berbahasa Indonesia dengan distribusi kelas yang sangat tidak seimbang, khususnya ketika kelas netral hanya mencapai 1,25% dari dataset. Kedua, penelitian menganalisis dampak ketidakseimbangan melalui evaluasi per kelas dan error analysis sehingga dapat diketahui mengapa accuracy tinggi tidak diikuti kemampuan mengenali kelas minoritas. Ketiga, penelitian memberikan implikasi praktis bagi pengembangan analitik umpan balik layanan kesehatan digital dengan menunjukkan bahwa informasi dari kelas negatif dan netral perlu tetap diperhatikan meskipun jumlahnya jauh lebih kecil. Kontribusi tersebut diharapkan memperkaya kajian penerapan TextCNN pada data kesehatan nyata sekaligus memberikan dasar untuk pengembangan strategi penyeimbangan data pada penelitian selanjutnya.

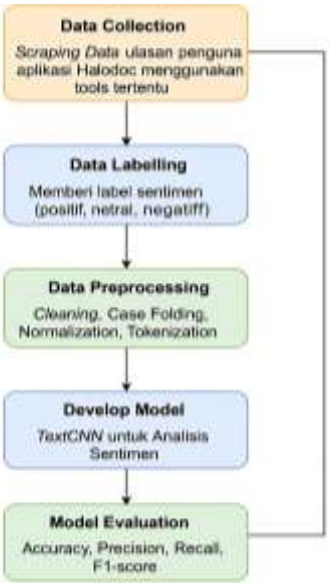
2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen dengan menerapkan model TextCNN pada dataset ulasan Halodoc. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, pelabelan data, pembangunan model, serta evaluasi performa. Dataset terdiri dari 2.000 ulasan yang dilabeli secara manual oleh satu annotator berdasarkan kriteria sentimen. Pendekatan ini berpotensi menimbulkan subjektivitas, sehingga menjadi salah satu keterbatasan dalam penelitian ini. Untuk mengatasi ketidakseimbangan data, digunakan teknik class weighting yang memberikan bobot lebih besar pada kelas minoritas. Model TextCNN dibangun dengan parameter tertentu yang telah ditentukan untuk memastikan proses pelatihan dapat direplikasi.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

2.1 Tahapan Penelitian

Tahapan penelitian pada studi ini ditunjukkan pada Gambar 1:



Gambar 1. Tahapan Penelitian

Dari Gambar 1, tahapan penelitian dapat dijelaskan sebagai berikut:

a. Pengumpulan Data
   Tahap awal penelitian adalah pengumpulan data berupa ulasan pengguna aplikasi Halodoc yang diambil dari Google Play Store. Proses pengambilan data dilakukan dengan teknik web scraping menggunakan Google Play Scraper sehingga diperoleh sebanyak 2.000 ulasan pengguna. Data yang dikumpulkan berupa teks ulasan yang merepresentasikan pengalaman dan opini pengguna terhadap layanan aplikasi. Data ulasan pengguna aplikasi Halodoc yang dikumpulkan pada periode 5 Juli 2025 hingga 6 Juli 2025. Rentang waktu tersebut dipilih untuk merepresentasikan ulasan pengguna yang relatif terbaru dan mencerminkan kondisi layanan aplikasi dalam periode operasional yang sama. Selama rentang waktu tersebut, diperoleh total 2.000 ulasan pengguna yang kemudian digunakan sebagai dataset penelitian. Penentuan rentang waktu pengambilan data bertujuan untuk memastikan konsistensi konteks layanan dan menghindari bias akibat perubahan fitur aplikasi yang signifikan di luar periode tersebut.
b. Pelabelan Sentimen
   Setelah data terkumpul, setiap ulasan dilabeli secara manual ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Pelabelan manual dilakukan untuk memastikan akurasi label sesuai dengan konteks ulasan. Hasil pelabelan menunjukkan distribusi sentimen yang tidak seimbang, dengan dominasi sentimen positif.
c. Preprocessing Teks
   Data teks yang telah dilabeli kemudian melalui tahap preprocessing untuk meningkatkan kualitas data dan mengurangi noise. Tahapan preprocessing meliputi proses lowercasing, penghapusan simbol dan angka, stopword removal, tokenisasi, dan padding. Preprocessing ini bertujuan untuk menyiapkan data agar dapat diproses secara optimal oleh model TextCNN.
d. Pembagian Dataset
   Dataset yang telah diproses selanjutnya dibagi menjadi data pelatihan dan data pengujian dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Pembagian data dilakukan secara acak dengan tetap mempertahankan proporsi kelas sentimen.
e. Pembangunan Arsitektur TextCNN
   Pada tahap ini, model TextCNN dibangun dengan beberapa lapisan utama, yaitu embedding layer, convolutional layer dengan beberapa ukuran kernel, max pooling layer, fully connected layer, dan softmax output layer. Arsitektur ini dirancang untuk mengekstraksi fitur lokal berupa pola n-gram dari teks ulasan.
f. Pelatihan Model
   Model dilatih menggunakan optimizer Adam dengan learning rate 0.001 selama 10 epoch. Panjang maksimum input ditetapkan sebesar 100 token dengan padding. Class weighting dihitung berdasarkan distribusi frekuensi kelas untuk mengurangi bias terhadap kelas mayoritas.
g. Evaluasi Model
   Setelah proses pelatihan selesai, model dievaluasi menggunakan data pengujian. Evaluasi dilakukan dengan menggunakan metrik accuracy, precision, recall, F1-score, dan confusion matrix. Evaluasi ini bertujuan untuk mengukur performa model secara keseluruhan serta pada masing-masing kelas sentimen.

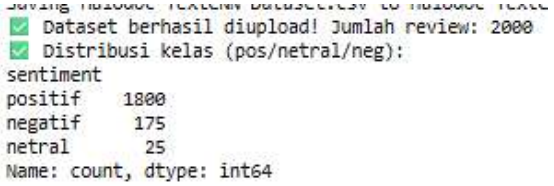
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

3. HASIL DAN PEMBAHASAN

Hasil analisis menunjukkan bahwa distribusi sentimen dalam dataset sangat tidak seimbang, dengan dominasi ulasan positif yang signifikan dibandingkan kelas lainnya. Kondisi ini berdampak langsung pada performa model. Model TextCNN menghasilkan akurasi yang tinggi, namun performa tersebut didominasi oleh keberhasilan dalam mengklasifikasikan kelas mayoritas. Hal ini terlihat dari nilai macro F1-score yang relatif rendah, yang mengindikasikan ketidakmampuan model dalam menangani kelas minoritas secara efektif. Confusion matrix menunjukkan bahwa model gagal mengidentifikasi kelas netral dengan nilai recall sebesar 0%. Hal ini mengindikasikan bahwa model tidak mampu mempelajari pola yang representatif dari kelas tersebut, yang kemungkinan disebabkan oleh jumlah data yang sangat terbatas. Selain itu, karakteristik ulasan yang cenderung pendek dan ambigu juga mempengaruhi kemampuan model dalam memahami konteks sentimen. Oleh karena itu, performa model tidak hanya dipengaruhi oleh arsitektur yang digunakan, tetapi juga oleh kualitas dan distribusi data.

Minimnya ulasan netral dapat disebabkan oleh perilaku pengguna yang cenderung memberikan ulasan hanya ketika memiliki pengalaman yang sangat positif atau sangat negatif. Hal ini menyebabkan distribusi data menjadi tidak representatif untuk pembelajaran model klasifikasi multikelas. Temuan ini menunjukkan bahwa penggunaan TextCNN tanpa strategi tambahan kurang memadai untuk menangani dataset dengan ketidakseimbangan ekstrem. Oleh karena itu, diperlukan pendekatan lain seperti data augmentation, resampling, atau penggunaan model berbasis transformer untuk meningkatkan performa pada kelas minoritas.

3.1 Distribusi Sentimen dan Implikasinya terhadap Pembelajaran Model



Gambar 2. Distribusi Sentimen

Distribusi sentimen dihitung berdasarkan jumlah ulasan pada masing-masing kelas dibandingkan dengan total data. Dari 2.000 ulasan, sebanyak 1.800 ulasan termasuk sentimen positif, 175 ulasan sentimen negatif, dan 25 ulasan sentimen netral. Persentase sentimen dihitung menggunakan perbandingan jumlah data per kelas terhadap total data, sehingga diperoleh distribusi 90% positif, 8,75% negatif, dan 1,25% netral. Hasil ini sesuai dengan visualisasi distribusi sentimen yang ditunjukkan pada Gambar 2.

Analisis distribusi sentimen menunjukkan bahwa dataset didominasi secara ekstrem oleh sentimen positif. Dari total 2.000 ulasan pengguna, sebanyak 1.800 ulasan (90%) diklasifikasikan sebagai sentimen positif, sementara 175 ulasan (8,75%) termasuk sentimen negatif dan hanya 25 ulasan (1,25%) yang bersifat netral. Pola distribusi ini mencerminkan karakteristik ulasan pengguna aplikasi Halodoc pada periode pengambilan data, di mana pengalaman positif lebih sering diekspresikan dibandingkan pengalaman netral atau negatif.

Distribusi kelas yang sangat tidak seimbang ini memiliki implikasi langsung terhadap proses pembelajaran model. Model deep learning cenderung memprioritaskan pola yang paling sering muncul selama pelatihan, sehingga berpotensi menghasilkan bias terhadap kelas mayoritas. Oleh karena itu, performa model tidak dapat dievaluasi hanya berdasarkan akurasi keseluruhan, melainkan perlu dianalisis menggunakan metrik yang mempertimbangkan ketimpangan antar kelas. Distribusi sentimen pada dataset menunjukkan ketidakseimbangan yang sangat signifikan, di mana ulasan positif mendominasi secara ekstrem dibandingkan kelas negatif dan netral. Kondisi ini berpotensi menyebabkan model mengalami bias terhadap kelas mayoritas. Minimnya jumlah data pada kelas netral juga menjadi faktor utama yang mempersulit model dalam mempelajari pola yang representatif.

3.2 Karakteristik Panjang Ulasan dan Tantangan Linguistik

Tabel 1. Profil Data

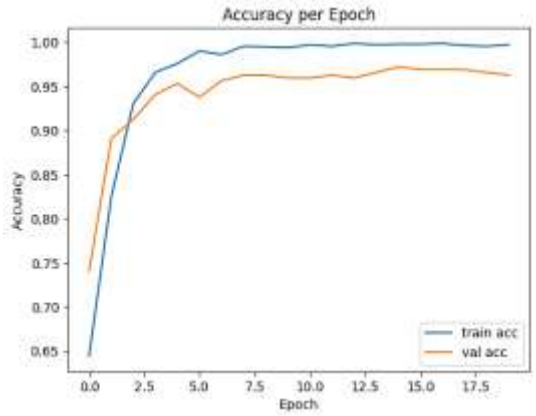
Statistik	Nilai	Penjelasan
Count	2000	Jumlah total ulasan dalam dataset ini adalah 2000.
Mean	8.06	Rata-rata panjang ulasan adalah sekitar 8 kata/karakter per ulasan.
Std	10.95	Standar deviasi (sebaran data) adalah sekitar 10.96. Nilai ini yang relatif tinggi dibandingkan mean menunjukkan bahwa ada variasi yang cukup besar dalam panjang ulasan.
Min	1	Panjang ulasan terpendek (minimal) adalah 1 kata/karakter.
25%	2	Kuartil pertama: 25% ulasan memiliki panjang 2 kata/karakter atau kurang.
50%	4	Median (nilai tengah): Setengah dari ulasan memiliki panjang 4 kata/karakter atau kurang.
75%	9	Kuartil ketiga: 75% ulasan memiliki panjang 9 kata/karakter atau kurang.
Max	87	Panjang ulasan terpanjang (maksimal) adalah 87 kata/karakter.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

Tabel 1 memperlihatkan distribusi panjang ulasan pengguna. Sebagian besar ulasan memiliki panjang antara 1 hingga 9 kata dengan rata-rata panjang ulasan sebesar 8 kata. Ulasan yang sangat pendek cenderung bersifat ambigu dan minim konteks, sehingga dapat menyulitkan model dalam membedakan sentimen, khususnya pada kelas netral. Analisis panjang ulasan menunjukkan bahwa sebagian besar ulasan pengguna bersifat sangat singkat. Rata-rata panjang ulasan adalah 8 kata, dengan median 4 kata. Sebagian besar ulasan berada pada rentang 1 hingga 9 kata, sementara hanya sebagian kecil ulasan yang memiliki panjang lebih dari 50 kata. Pola ini menunjukkan bahwa pengguna cenderung menyampaikan opini secara ringkas dan langsung. Ulasan dengan panjang yang pendek sering kali minim konteks dan bersifat ambigu. Kata-kata seperti “bagus”, “oke”, atau “lumayan” dapat merepresentasikan sentimen positif maupun netral, tergantung pada konteks yang tidak selalu tersurat dalam teks. Kondisi ini menjadi tantangan tersendiri bagi model klasifikasi sentimen, terutama pada kelas netral yang secara semantik berada di antara sentimen positif dan negatif.

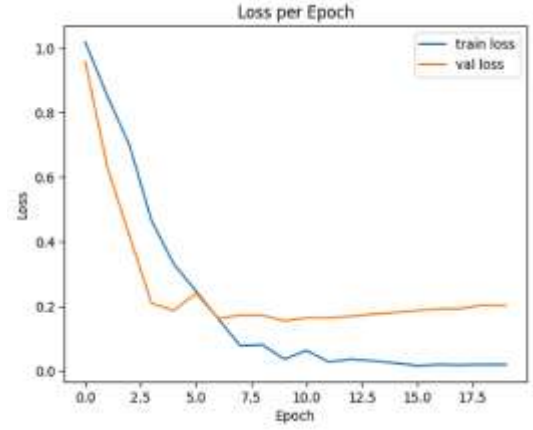
Dalam konteks arsitektur TextCNN, keterbatasan ini semakin terlihat karena model mengandalkan pola lokal berbasis n-gram. Ketika teks input sangat pendek, jumlah kombinasi n-gram yang dapat dipelajari menjadi terbatas, sehingga representasi fitur yang dihasilkan kurang mampu menangkap nuansa sentimen secara menyeluruh. Ulasan dengan panjang yang sangat pendek cenderung mengandung informasi yang terbatas, sehingga menyulitkan model dalam menangkap konteks sentimen secara akurat. Hal ini menjadi salah satu faktor yang mempengaruhi performa model.

3.3 Hasil Pelatihan Model TextCNN



Gambar 3. Validation Accuracy

Gambar 3 menunjukkan kurva akurasi validasi selama proses pelatihan model TextCNN. Terlihat bahwa akurasi validasi meningkat secara signifikan pada epoch awal dan mulai stabil pada epoch menengah hingga akhir, dengan nilai maksimum mencapai 93,75%. Pola ini menunjukkan bahwa model yang didominasi oleh kelas mayoritas sehingga belum mampu melakukan generalisasi dengan baik terhadap data minoritas. Selama proses pelatihan, model TextCNN menunjukkan peningkatan performa yang konsisten. Akurasi pelatihan meningkat dari 75,5% pada epoch awal menjadi 99,8% pada epoch terakhir. Sementara itu, akurasi validasi mencapai nilai maksimum sebesar 93,75%. Peningkatan ini menunjukkan bahwa model mampu mempelajari pola sentimen dari data pelatihan dan mempertahankan performa yang relatif stabil pada data validasi.



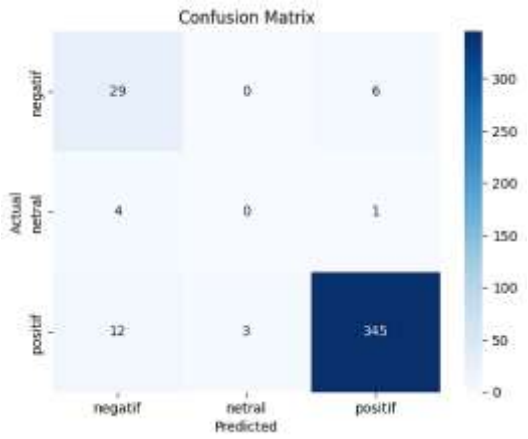
Gambar 4. Training Loss

Gambar 4 memperlihatkan penurunan nilai loss pelatihan dan validasi secara bertahap seiring bertambahnya epoch. Penurunan loss yang konsisten menunjukkan bahwa proses optimasi berjalan dengan baik,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

meskipun terdapat fluktuasi kecil pada loss validasi yang mengindikasikan potensi overfitting ringan. Nilai loss pelatihan dan validasi mengalami penurunan yang gradual sepanjang epoch, meskipun terdapat fluktuasi kecil pada loss validasi di epoch akhir. Fluktuasi ini mengindikasikan adanya kecenderungan overfitting ringan, namun masih berada dalam batas yang dapat diterima. Penerapan teknik class weighting berperan dalam mengurangi dominasi kelas mayoritas selama proses pembelajaran, meskipun belum sepenuhnya mengatasi ketimpangan distribusi kelas. Meskipun teknik class weighting telah diterapkan, hasil menunjukkan bahwa metode ini belum mampu mengatasi ketidakseimbangan data secara efektif, terutama pada kelas netral. Hal ini terlihat dari nilai recall yang tetap rendah, sehingga diperlukan pendekatan tambahan seperti oversampling atau data augmentation.

3.4 Evaluasi Model Menggunakan Confusion Matrix



Gambar 5. Confusion Matrix

Gambar 5 menunjukkan confusion matrix hasil klasifikasi sentimen. Model mampu mengklasifikasikan sentimen positif dengan baik, namun performa pada kelas netral tidak dapat mengidentifikasi ulasan netral. Hal ini disebabkan oleh jumlah data netral yang sangat terbatas dibandingkan kelas lainnya. Confusion matrix memberikan gambaran yang lebih rinci mengenai performa model pada masing-masing kelas sentimen. Model menunjukkan kemampuan yang sangat baik dalam mengklasifikasikan sentimen positif, dengan tingkat recall sebesar 96%. Hal ini sejalan dengan dominasi kelas positif dalam dataset, yang memungkinkan model mempelajari pola sentimen positif secara lebih representatif.

Sebaliknya, performa pada kelas netral dengan nilai recall sebesar 0%. Seluruh ulasan netral pada data uji diklasifikasikan ke dalam kelas lain, terutama kelas positif. Fenomena ini menunjukkan bahwa model kesulitan membedakan sentimen netral dari sentimen positif, yang disebabkan oleh kemiripan kosakata dan keterbatasan konteks pada ulasan netral yang umumnya sangat singkat. Kesalahan klasifikasi juga ditemukan pada kelas negatif, meskipun dalam jumlah yang lebih terbatas. Beberapa ulasan negatif diklasifikasikan sebagai positif karena mengandung kata-kata yang secara leksikal bernuansa positif, namun berada dalam konteks keluhan atau kritik implisit. Hal ini menunjukkan bahwa model berbasis n-gram masih memiliki keterbatasan dalam menangkap hubungan semantik yang lebih kompleks.

Berdasarkan confusion matrix, model gagal mengklasifikasikan seluruh data pada kelas netral dengan nilai recall sebesar 0%. Hal ini menunjukkan kegagalan model dalam mengenali salah satu kelas sentimen, sehingga performa model tidak dapat dikatakan optimal dalam tugas klasifikasi multikelas.

3.5 Analisis Precision, Recall, dan F1-score

Tabel 2. Classification Report

Kelas	Precision	Recall	F1-score	Support
Negatif	0,64	0,83	0,72	35
Netral	0,00	0,00	0,00	5
Positif	0,98	0,96	0,97	360
Accuracy			0,94	400
Macro Avg	0,54	0,60	0,56	400
Weighted Avg	0,94	0,94	0,94	400

Tabel 2 menunjukkan nilai precision, recall, dan F1-score untuk masing-masing kelas sentimen. Nilai weighted F1-score sebesar 0,94 menunjukkan performa model yang baik secara keseluruhan, namun nilai macro F1-score sebesar 0,57 mengindikasikan ketimpangan performa antar kelas.

Hasil evaluasi menggunakan classification report menunjukkan bahwa weighted F1-score mencapai nilai 0,94, yang mengindikasikan performa model cukup baik secara keseluruhan. Namun, nilai macro F1-score hanya

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

sebesar 0,57, mencerminkan ketimpangan performa antar kelas sentimen. Meskipun model menghasilkan nilai akurasi yang tinggi, metrik tersebut tidak sepenuhnya merepresentasikan performa model secara keseluruhan karena dipengaruhi oleh dominasi kelas mayoritas. Hal ini terlihat dari nilai macro F1-score yang lebih rendah, yang menunjukkan bahwa model belum mampu bekerja secara seimbang pada seluruh kelas.

3.6 Analisis Kesalahan (Error Analysis)

Analisis terhadap ulasan yang salah diklasifikasikan menunjukkan bahwa sebagian besar kesalahan terjadi pada ulasan yang pendek dan ambigu. Ulasan netral sering kali mengandung kata-kata yang juga muncul pada ulasan positif, sehingga model cenderung memprediksinya sebagai sentimen positif. Selain itu, beberapa ulasan negatif memiliki struktur kalimat yang tidak eksplisit atau bercampur bahasa, yang menyulitkan model dalam memahami konteks sentimen secara akurat. Temuan ini menunjukkan bahwa meskipun TextCNN efektif dalam menangkap pola lokal pada teks, performanya tetap sangat dipengaruhi oleh kualitas representasi konteks dan distribusi data. Pada kondisi data nyata yang tidak seimbang, keterbatasan ini menjadi semakin signifikan.

3.7 Implikasi Ketidakseimbangan Data pada Analisis Sentimen Aplikasi Kesehatan

Ketidakseimbangan distribusi sentimen yang ekstrem pada dataset ulasan Halodoc memiliki implikasi yang signifikan terhadap interpretasi hasil analisis sentimen. Dalam konteks aplikasi layanan kesehatan digital, dominasi ulasan positif tidak hanya memengaruhi proses pembelajaran model, tetapi juga berpotensi membentuk persepsi yang bias terhadap kualitas layanan apabila hasil analisis digunakan sebagai dasar pengambilan keputusan.

Model klasifikasi sentimen yang dilatih pada data yang sangat tidak seimbang cenderung mengoptimalkan performa pada kelas mayoritas, sehingga kesalahan pada kelas minoritas menjadi kurang terdeteksi ketika hanya mengandalkan metrik evaluasi global seperti akurasi. Pada penelitian ini, meskipun akurasi validasi mencapai 93,75%, performa pada kelas netral sangat rendah. Kondisi ini menunjukkan bahwa hasil analisis sentimen perlu ditafsirkan secara hati-hati, terutama apabila digunakan untuk mengevaluasi kepuasan pengguna secara menyeluruh. Dalam praktiknya, ulasan netral dapat merepresentasikan pengalaman pengguna yang berada di batas antara kepuasan dan ketidakpuasan. Kegagalan model dalam mengidentifikasi sentimen netral berpotensi menyebabkan informasi penting terkait aspek layanan yang perlu ditingkatkan menjadi terabaikan. Oleh karena itu, penelitian ini menegaskan pentingnya mempertimbangkan distribusi kelas dan karakteristik data dalam penerapan sistem analisis sentimen pada domain layanan kesehatan.

3.8 Keterbatasan Arsitektur TextCNN pada Data Ulasan Pendek

Arsitektur TextCNN dikenal efektif dalam menangkap pola lokal teks melalui operasi konvolusi berbasis n-gram. Keunggulan ini memungkinkan model untuk mengenali frasa-frasa kunci yang sering muncul pada ulasan positif maupun negatif. Namun, pada dataset dengan karakteristik ulasan yang sangat pendek, seperti pada penelitian ini, efektivitas TextCNN menjadi terbatas. Ulasan dengan panjang rata-rata delapan kata menyediakan konteks yang sangat terbatas bagi model untuk membentuk representasi sentimen yang kaya. Ketika ulasan hanya terdiri dari satu atau dua kata, pola n-gram yang dapat dipelajari menjadi sangat minim. Akibatnya, model lebih bergantung pada kemunculan kata tertentu tanpa mempertimbangkan hubungan semantik yang lebih luas. Keterbatasan ini terlihat jelas pada kelas netral, di mana kosakata yang digunakan sering kali tumpang tindih dengan kelas positif maupun negatif. TextCNN yang berfokus pada pola lokal tidak selalu mampu membedakan nuansa sentimen yang bersifat implisit atau ambigu. Temuan ini menunjukkan bahwa meskipun TextCNN memiliki keunggulan komputasi dan kesederhanaan arsitektur, pendekatan ini masih memerlukan pengayaan konteks atau integrasi dengan model lain untuk menangani data ulasan yang pendek dan ambigu secara lebih efektif.

3.9 Justifikasi Pemilihan TextCNN dan Skema Klasifikasi Tiga Kelas

Pemilihan arsitektur TextCNN dalam penelitian ini didasarkan pada kemampuannya dalam memproses data teks secara efisien serta keberhasilannya pada berbagai studi analisis sentimen sebelumnya. TextCNN memiliki struktur yang relatif sederhana dibandingkan model berbasis transformer, sehingga lebih mudah diimplementasikan dan diinterpretasikan, terutama pada dataset berukuran menengah seperti ulasan aplikasi. Penggunaan skema klasifikasi tiga kelas sentimen, yaitu positif, negatif, dan netral, bertujuan untuk merepresentasikan spektrum opini pengguna secara lebih realistis. Kelas netral memiliki peran penting dalam menggambarkan pengalaman pengguna yang tidak sepenuhnya puas maupun tidak puas. Meskipun performa model pada kelas ini masih rendah, keberadaan kelas netral tetap relevan secara konseptual dalam analisis sentimen layanan kesehatan.

Selain itu, penerapan teknik class weighting dipilih sebagai upaya untuk mengurangi bias pembelajaran akibat ketidakseimbangan data. Teknik ini memungkinkan model memberikan bobot kesalahan yang lebih besar pada kelas minoritas selama proses pelatihan. Meskipun belum sepenuhnya mengatasi permasalahan performa pada kelas netral, pendekatan ini terbukti membantu meningkatkan stabilitas pelatihan dan menjaga performa pada kelas negatif.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

Meskipun teknik class weighting telah diterapkan, hasil menunjukkan bahwa metode ini belum mampu mengatasi ketidakseimbangan data secara efektif, terutama pada kelas netral. Hal ini terlihat dari nilai recall yang tetap rendah, sehingga diperlukan pendekatan tambahan seperti oversampling atau data augmentation.

3.10 Implikasi Praktis bagi Pengembangan Layanan Kesehatan Digital

Hasil penelitian ini memiliki implikasi praktis bagi pengembang aplikasi layanan kesehatan digital seperti Halodoc. Analisis sentimen berbasis TextCNN dapat digunakan sebagai alat bantu untuk memantau persepsi pengguna terhadap layanan secara cepat dan otomatis. Namun, hasil analisis perlu dikombinasikan dengan pendekatan lain, seperti analisis tematik atau evaluasi manual, untuk memastikan bahwa masukan dari kelas minoritas tidak terabaikan. Dalam konteks peningkatan kualitas layanan, ulasan negatif dan netral sering kali mengandung informasi yang lebih bernilai dibandingkan ulasan positif. Oleh karena itu, sistem analisis sentimen sebaiknya dirancang untuk secara khusus menyoroti ulasan-ulasan tersebut, meskipun jumlahnya relatif sedikit. Penelitian ini menekankan bahwa performa model yang tinggi pada kelas mayoritas tidak serta-merta menjamin efektivitas sistem dalam mendukung pengambilan keputusan strategis.

3.11 Pembahasan

Hasil penelitian menunjukkan bahwa TextCNN mampu mencapai accuracy sebesar 93,75% dan weighted F1-score sebesar 0,94, tetapi macro F1-score hanya 0,57 dan recall kelas netral sebesar 0%. Temuan ini memperlihatkan bahwa kinerja agregat model sangat dipengaruhi oleh dominasi kelas positif. Dengan kata lain, accuracy 93,75% tidak dapat ditafsirkan sebagai kemampuan model yang merata pada ketiga kelas. Hasil tersebut konsisten dengan analisis confusion matrix dan error analysis yang menunjukkan bahwa sebagian besar kesalahan terjadi ketika ulasan netral atau negatif memiliki kosakata yang mirip dengan kelas positif atau ketika teks terlalu pendek untuk menyediakan konteks yang cukup.

Jika dibandingkan dengan penelitian Sidabutar dan Juardi (2025), penelitian tersebut menggunakan 5.000 ulasan Halodoc dan SVM untuk klasifikasi positif-negatif serta melaporkan akurasi tinggi [4]. Penelitian ini menggunakan objek Halodoc yang sama tetapi memiliki tiga kelas sentimen dan fokus yang lebih kuat pada ketidakseimbangan ekstrem. Karena dataset, jumlah kelas, dan model berbeda, nilai accuracy tidak dapat dibandingkan secara langsung. Namun, perbandingan tersebut menunjukkan bahwa pemilihan model dan skema label perlu disesuaikan dengan tujuan analisis. Pada penelitian ini, penambahan kelas netral memperlihatkan persoalan yang tidak terlihat ketika klasifikasi hanya menggunakan dua kelas.

Dibandingkan dengan Imaduddin et al. (2023), yang menggunakan IndoBERT pada 9.310 ulasan dari Halodoc, Alodokter, dan KlikDokter dan memperoleh F1-score 95% [9], hasil TextCNN pada penelitian ini lebih rendah pada metrik yang mempertimbangkan keseimbangan kelas. Perbedaan tersebut tidak berarti TextCNN tidak efektif, karena IndoBERT memiliki representasi kontekstual yang lebih kuat dan dataset Imaduddin et al. memiliki skenario klasifikasi yang berbeda. Justru, hasil penelitian ini memperlihatkan bahwa pada dataset yang sangat tidak seimbang, pemilihan metrik dan distribusi kelas memiliki pengaruh besar terhadap interpretasi performa.

Temuan penelitian ini juga dapat dibandingkan dengan Al-Hadhrami et al. (2024), yang memperoleh accuracy hingga 96% menggunakan Bi-LSTM-CNN pada ulasan obat pasien [6]. Model hibrida tersebut memanfaatkan informasi sekuensial dan konvolusional, sedangkan TextCNN pada penelitian ini berfokus pada pola lokal. Perbedaan domain dan dataset membuat perbandingan angka tidak bersifat langsung, tetapi hasil keduanya menunjukkan bahwa arsitektur deep learning dapat efektif pada teks kesehatan apabila karakteristik data mendukung. Sementara itu, Ouni et al. (2024) menunjukkan bahwa pengayaan representasi kata melalui SoftEMB dapat meningkatkan performa beberapa model deep learning [7]. Temuan tersebut sejalan dengan hasil penelitian ini yang menunjukkan bahwa representasi fitur menjadi faktor penting ketika teks pendek dan ambigu.

Penelitian Li et al. (2025) memberikan pembanding yang sangat relevan karena juga menangani data teks kesehatan yang tidak seimbang. Li et al. menggunakan SMOTE untuk menyeimbangkan data dan memperoleh performa SVM yang baik pada data eksternal [3]. Pada penelitian ini, class weighting belum mampu mengatasi kegagalan model pada kelas netral. Perbedaan tersebut menunjukkan bahwa pembobotan kelas saja mungkin belum cukup ketika jumlah sampel minoritas sangat kecil, yaitu hanya 25 dari 2.000 data. Oleh karena itu, oversampling, data augmentation, atau kombinasi strategi penyeimbangan dapat menjadi arah pengembangan yang logis. Penelitian Ye et al. (2025) juga menunjukkan bahwa TextCNN dapat diperkuat dengan representasi konteks global [8], yang relevan dengan keterbatasan penelitian ini pada ulasan pendek.

Secara keseluruhan, hasil penelitian menempatkan TextCNN sebagai model yang mampu menghasilkan performa tinggi pada kelas mayoritas tetapi belum cukup robust terhadap ketidakseimbangan ekstrem. Temuan utama bukan hanya accuracy 93,75%, melainkan kesenjangan antara weighted F1-score 0,94 dan macro F1-score 0,57 serta recall netral 0%. Perbandingan dengan penelitian terdahulu memperkuat bahwa keberhasilan model klasifikasi sentimen perlu dinilai bersama karakteristik dataset, jumlah kelas, distribusi label, representasi teks, dan metrik per kelas. Dalam konteks layanan kesehatan digital, kegagalan mengenali kelas minoritas perlu diperhatikan karena ulasan negatif dan netral dapat berisi informasi yang penting untuk perbaikan layanan. Oleh sebab itu, penelitian ini memberikan bukti bahwa penggunaan TextCNN dengan class weighting merupakan langkah awal yang layak, tetapi belum memadai untuk distribusi kelas yang sangat ekstrem.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

4. KESIMPULAN

Berdasarkan hasil penelitian, model TextCNN mampu mencapai accuracy sebesar 93,75% dan weighted F1-score sebesar 0,94 pada 2.000 ulasan Halodoc berbahasa Indonesia. Namun, macro F1-score hanya sebesar 0,57 dan recall kelas netral sebesar 0%, yang menunjukkan bahwa performa model tidak merata pada seluruh kelas. Dominasi 90% ulasan positif menyebabkan model lebih mudah mempelajari pola kelas mayoritas, sedangkan hanya 25 ulasan netral membuat pola kelas tersebut tidak cukup representatif untuk dipelajari. Dengan demikian, accuracy tinggi tidak dapat dijadikan satu-satunya dasar untuk menyimpulkan bahwa model telah bekerja optimal. Kontribusi penelitian ini terdiri atas tiga hal. Pertama, penelitian memberikan evaluasi empiris TextCNN pada ulasan Halodoc dengan tiga kelas sentimen dan ketidakseimbangan kelas yang ekstrem. Kedua, penelitian memperlihatkan melalui confusion matrix, classification report, dan error analysis bahwa class weighting belum mampu mengatasi kegagalan model pada kelas netral. Ketiga, penelitian memberikan implikasi praktis bahwa sistem analisis sentimen layanan kesehatan perlu mempertahankan perhatian terhadap ulasan negatif dan netral meskipun jumlahnya jauh lebih sedikit. Hasil perbandingan penelitian terdahulu menunjukkan bahwa perbedaan model, representasi teks, skema kelas, dan distribusi data dapat menghasilkan karakteristik performa yang berbeda, sehingga evaluasi harus mempertimbangkan konteks dataset. Untuk penelitian selanjutnya, disarankan menerapkan oversampling atau data augmentation pada kelas minoritas, melakukan pengujian strategi penyeimbangan secara sistematis, serta mempertimbangkan model hibrida atau pendekatan berbasis transformer yang dapat memberikan representasi konteks lebih kaya. Dengan strategi tersebut, kemampuan model dalam mengenali sentimen negatif dan terutama netral diharapkan dapat ditingkatkan sehingga hasil analisis lebih representatif untuk mendukung peningkatan kualitas layanan kesehatan digital.

REFERENCES

[1] F. Binsar, Mts. Arief, V. U. Tjhin, and I. Susilowati, "Exploring consumer sentiments in telemedicine and telehealth services: Towards an integrated framework for innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 11, no. 1, p. 100453, 2025, doi: https://doi.org/10.1016/j.joitmc.2024.100453.

[2] A. Jerfy, O. Selden, and R. Balkrishnan, "The Growing Impact of Natural Language Processing in Healthcare and Public Health," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, vol. 61, p. 469580241290095, Aug. 2024, doi: 10.1177/00469580241290095.

[3] Q. and K. C. and W. J. and L. G. and F. X. and L. X. and Y. G. Li Xiadong and Shu, "An Intelligent System for Classifying Patient Complaints Using Machine Learning and Natural Language Processing: Development and Validation Study," *J Med Internet Res*, vol. 27, p. e55721, Jan. 2025, doi: 10.2196/55721.

[4] T. G. W. M. Sidabutar and D. Juardi, "Analisis Sentimen Masyarakat Terhadap Penggunaan Halodoc Sebagai Layanan Telemedicine Di Indonesia," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 1, Jan. 2025, doi: 10.23960/jitet.v13i1.5682.

[5] H. Imaduddin, F. Yusfida A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach." [Online]. Available: www.ijacsa.thesai.org

[6] S. Al-Hadhrami, T. Vinko, T. Al-Hadhrami, F. Saeed, and S. Qasem, "Deep learning-based method for sentiment analysis for patients' drug reviews," *PeerJ Comput. Sci.*, vol. 10, p. e1976, Aug. 2024, doi: 10.7717/peerj-cs.1976.

[7] C. Ouni, E. Benmohamed, and H. Ltifi, "Deep learning-based Soft word embedding approach for sentiment analysis," *Procedia Comput. Sci.*, vol. 246, pp. 1355–1364, 2024, doi: https://doi.org/10.1016/j.procs.2024.09.720.

[8] Y. Ye, X. Chang, and A. Fan, "NE ZHA-TextCNN Method for Multi-Label Long Text Classification," *Procedia Comput. Sci.*, vol. 262, pp. 313–319, 2025, doi: https://doi.org/10.1016/j.procs.2025.05.058.

[9] H. Imaduddin, F. A'la, and Y. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, Jan. 2023, doi: 10.14569/IJACSA.2023.0140813.

[10] F. Gräßer, H. Malberg, S. Kallumadi, and S. Zaunseder, "Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning," ACM International Conference Proceeding Series, 2018, pp. 121–125, doi: 10.1145/3194658.3194677.

[11] X. Jiang, C. Song, Y. Xu, Y. Li, and Y. Peng, "Research on Sentiment Classification for Netizens Based on the BERT-BiLSTM-TextCNN Model," PeerJ Computer Science, vol. 8, 2022, e1005, doi: 10.7717/peerj-cs.1005.

[12] S. Natu and M. Aparicio, "Analyzing Knowledge Sharing Behaviors in Virtual Teams: Practical Evidence from Digitalized Workplaces," Journal of Innovation and Knowledge, vol. 7, no. 4, 2022, article 100248, doi: 10.1016/j.jik.2022.100248

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 4, August 2026, Page 150-159
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i4.9331

[13] T. G. W. M. Sidabutar and D. Juardi, "Analisis Sentimen Masyarakat Terhadap Penggunaan Halodoc sebagai Layanan Telemedicine di Indonesia," Jurnal Informatika dan Teknik Elektro Terapan, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5682.

[14] R. Stewart, J. Chaturvedi, and A. Roberts, "Natural Language Processing—Relevance to Patient Outcomes and Real-World Evidence," Expert Review of Pharmacoeconomics & Outcomes Research, vol. 24, no. 1, 2024, pp. 5–9, doi: 10.1080/14737167.2023.2275670.

[15] A. Rawat, M. A. Wani, M. ElAffendi, A. S. Imran, Z. Kastrati, and S. M. Daudpota, "Drug Adverse Event Detection Using Text-Based Convolutional Neural Networks (TextCNN) Technique," Electronics (Switzerland), vol. 11, no. 20, pp. 1–21, 2022, doi: 10.3390/electronics11203336.

[16] K. Denecke and D. Reichenpfader, "Sentiment Analysis of Clinical Narratives: A Scoping Review," Journal of Biomedical Informatics, vol. 140, 2023, article 104336, doi: 10.1016/j.jbi.2023.104336.

[17] P. K. Nag, A. Bhagat, R. Vishnu Priya, and D. K. Khare, "Emotional Intelligence Through Artificial Intelligence: NLP and Deep Learning in the Analysis of Healthcare Texts," in 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), 2023, pp. 1–7, doi: 10.1109/ICAIIHI57871.2023.10489117.

[18] Y. Ling, "Bio+Clinical BERT, BERT Base, and CNN Performance Comparison for Predicting Drug-Review Satisfaction," arXiv preprint arXiv:2308.03782, 2023, doi: 10.48550/arXiv.2308.03782.

[19] A. Ghosh, S. Umer, B. C. Dhara, and G. G. M. N. Ali, "A Multimodal Pain Sentiment Analysis System Using Ensembled Deep Learning Approaches for IoT-Enabled Healthcare Framework," Sensors, vol. 25, no. 4, 2025, article 1223, doi: 10.3390/s25041223.

[20] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify," STAIN (Seminar Nasional Teknologi & Sains), vol. 3, no. 1, 2024, pp. 145–153, doi: 10.29407/stains.v3i1.4166.

[21] M. Prasetya, M. Wulandari, and S. A. Nikmah, "Implementasi NLP (Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify," Stains (Seminar Nasional Teknologi & Sains), vol. 3, no. 1, pp. 145–153, 2024.

[22] D. Suhartono, K. Purwandari, N. H. Jeremy, S. Philip, P. Arisaputra, and I. H. Parmonangan, "Deep Neural Networks and Weighted Word Embeddings for Sentiment Analysis of Drug Product Reviews," Procedia Computer Science, vol. 216, 2023, pp. 664–671, doi: 10.1016/j.procs.2022.12.182.