

# PENERAPAN ALGORITMA BI-LSTM DENGAN OPTIMASI THRESHOLD ADJUSTMENT UNTUK ANALISIS SENTIMEN ULASAN APLIKASI MOBILE JKN

Adam Malik<sup>1\*</sup>, Bambang Irawan<sup>2</sup>

<sup>1,2</sup>Universitas Muhamadiyah Setiabudi; Jalan P. Diponegoro KM 2, Pesantunan, Wanasari, Brebes, Jawa Tengah 52212; Telp. (0283) 6199000 Fax: (0283) 6199001

---

**Keywords:**

Analisis Sentimen, Bi-LSTM, Threshold Adjustment, Mobile JKN, Deep Learning.

**Corespondent Email:**

adampfans61@gmail.com

**Abstrak.** Penelitian ini bertujuan untuk menerapkan algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) dengan optimasi threshold adjustment dalam analisis sentimen ulasan pengguna aplikasi Mobile JKN pada platform Google Play Store. Data ulasan yang digunakan berasal dari file mobilejkn.csv dengan ribuan record, diproses melalui tahapan pre-processing yang mencakup pembersihan teks, tokenisasi, penghapusan stopword, dan stemming. Model memanfaatkan lapisan embedding, bidirectional LSTM, dropout, serta dense layer dengan aktivasi softmax. Evaluasi model Bi-LSTM mencapai akurasi 88,5% pada data validasi (setelah pelatihan 10 epoch dengan optimizer Adam), dengan peningkatan performa menjadi 90,2% setelah penerapan threshold adjustment (penyesuaian batas probabilitas maksimum  $<0,58$  untuk klasifikasi netral). Nilai presisi rata-rata 89,1%, recall 88,7%, dan F1-score 88,9%. Hasil analisis menunjukkan dominasi sentimen negatif (sekitar 45-50%) terkait masalah teknis seperti kesulitan login, verifikasi OTP lambat, kegagalan booking antrian, serta proses registrasi yang rumit. Temuan ini sejalan dengan keluhan umum pada ulasan terbaru (rating rata-rata 4,3 dari 933 ribu ulasan). Penelitian ini merekomendasikan kepada BPJS Kesehatan untuk segera memperbaiki fitur autentikasi, stabilitas server, dan antarmuka pengguna agar meningkatkan kepuasan serta loyalitas peserta JKN.



Copyright © [JITET](#) (Jurnal Informatika dan Teknik Elektro Terapan). This article is an open access article distributed under terms and conditions of the Creative Commons Attribution (CC BY NC)

**Abstract.** This research aims to implement the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm with threshold adjustment optimization for sentiment analysis of Mobile JKN application user reviews on the Google Play Store platform. The review data utilized comes from the mobilejkn.csv file containing thousands of records, processed through pre-processing stages including text cleaning, tokenization, stopword removal, and stemming. The model employs embedding layers, bidirectional LSTM, dropout, and a dense layer with softmax activation. The Bi-LSTM model evaluation achieves an accuracy of 88.5% on validation data (after 10 epochs of training with Adam optimizer), with performance improvement to 90.2% after applying threshold adjustment (adjusting the maximum probability threshold  $<0.58$  for neutral classification). Average precision reaches 89.1%, recall 88.7%, and F1-score 88.9%. The analysis results indicate dominance of negative sentiment (approximately 45-50%) related to technical issues such as login difficulties, slow OTP verification, queue booking failures, and complicated registration processes. These findings align with common complaints in recent reviews (average rating 4.3 from 933 thousand reviews). This study recommends that BPJS Kesehatan immediately improve authentication features, server stability, and user interface to enhance participant satisfaction and loyalty.

## 1. PENDAHULUAN

Aplikasi Mobile JKN yang dikembangkan oleh BPJS Kesehatan merupakan instrumen digital utama untuk memfasilitasi akses layanan Jaminan Kesehatan Nasional (JKN) di Indonesia. Namun, meskipun rating rata-rata mencapai 4,3 bintang dari 933 ribu ulasan (per Januari 2026), terdapat tren penurunan kepuasan akibat keluhan berulang terkait performa teknis. Analisis sentimen menjadi pendekatan sistematis untuk mengidentifikasi pola opini pengguna secara empiris[1], sehingga memberikan dasar bagi pengembangan dalam pengambilan keputusan berbasis data.

Penelitian terdahulu menunjukkan efektivitas model deep learning pada ulasan aplikasi serupa. Penerapan IndoBERT pada ulasan Mobile JKN mampu mencapai akurasi hingga 97%[2] dalam klasifikasi sentimen, namun memerlukan sumber daya komputasi yang tinggi. Sebaliknya, metode klasik seperti Naïve Bayes dan Support Vector Machine (SVM) memiliki akurasi yang lebih rendah dan bergantung pada fitur statistik sehingga kurang robust terhadap teks informal. Model Bidirectional Long Short-Term Memory (Bi-LSTM) unggul dalam menangkap konteks dua arah[3] pada teks bahasa Indonesia yang kaya variasi slang dan dependensi temporal. Kesenjangan utama terletak pada minimnya optimasi threshold untuk mengatasi ambiguitas kelas netral pada dataset ulasan aplikasi kesehatan yang tidak seimbang.

Tujuan utama penelitian ini adalah: (1) membangun dan mengevaluasi model Bi-LSTM untuk klasifikasi sentimen ulasan Mobile JKN, (2) mengoptimasi performa melalui threshold adjustment, serta (3) memberikan rekomendasi perbaikan aplikasi berbasis temuan empiris.

## 2. TINJAUAN PUSTAKA

### 2.1. Analisis Sentimen

Analisis merupakan pendekatan metodologis dalam bidang pemrosesan data teks, dengan tujuan memunculkan pendapat dari pernyataan yang diberikan, kemudian menafsirkannya dalam kerangka kontekstual yang sesuai, dan mengevaluasinya dengan cara yang tidak memihak [4]. Prosedur ini diterapkan untuk menjelaskan informasi yang tertanam dalam opini, sehingga memfasilitasi

pemahaman yang lebih komprehensif tentang isi dan lintasan pendapat yang diartikulasikan.

### 2.2. Ulasan

Setelah menggunakan suatu barang atau layanan, konsumen melakukan evaluasi atau penilaian. Ulasan ini dapat membantu perusahaan memahami kebutuhan. Ulasan pengguna juga mencerminkan persepsi kualitas layanan secara langsung dan kepuasan pelanggan[5]. Selain itu, kritik dan evaluasi teks tertentu, baik fiksi maupun nonfiksi, termasuk novel, cerpen, puisi, film, musik, game, dan sebagainya, didasarkan pada pengamatan, pertimbangan, dan penelitian.

### 2.3. Aplikasi Mobile JKN

Mobile JKN adalah aplikasi mobile yang dikembangkan oleh BPJS Kesehatan untuk memfasilitasi akses layanan Jaminan Kesehatan Nasional (JKN) bagi peserta. Aplikasi ini menyediakan berbagai fitur seperti informasi kepesertaan, lokasi fasilitas kesehatan, pembayaran iuran, pendaftaran peserta baru, serta antrian online. Dengan pertumbuhan jumlah kepesertaan JKN yang mencapai lebih dari 222 juta jiwa, aplikasi ini menjadi instrumen utama dalam mendukung digitalisasi layanan kesehatan di Indonesia. Namun, berdasarkan data terbaru (Januari 2026), aplikasi Mobile JKN memiliki rata-rata rating 4.2 bintang dari 939 ribu ulasan dan lebih dari 50 juta unduhan di Google Play Store, dengan keluhan dominan terkait verifikasi lambat (SMS/OTP), kegagalan booking antrian, serta proses registrasi yang rumit.

### 2.4. Data Set

Dataset merupakan kumpulan entitas data bersama dengan atribut yang sesuai, di mana setiap entitas menggambarkan kasus atau contoh yang berbeda yang ditentukan oleh serangkaian karakteristik atau fitur yang dapat diukur. Kumpulan data tersebut dapat diklasifikasikan sebagai numerik, kategoris, atau hibrida dari keduanya, dan mereka memainkan peran penting sebagai input dasar untuk upaya analitis dan kegiatan data mining.

### 2.5 Text Mining

Text mining merupakan ekstraksi sistematis informasi terkait dari kumpulan dokumen melalui keterlibatan pengguna dengan

instrumen analitis, yang merupakan komponen integral dari proses data mining [6]. Selain itu, teks minning berkontribusi signifikan terhadap penanganan data tekstual yang tidak terstruktur. Tujuan utama penambangan teks adalah untuk memperoleh wawasan berharga dari konten tekstual yang tidak terstruktur atau hanya sedikit semi-terstruktur.

## 2.6. Algoritma Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional Long Short-Term Memory (Bi-LSTM) merupakan pengembangan dari arsitektur LSTM yang mampu memproses informasi secara dua arah (forward and backward). Hal ini memungkinkan model untuk menangkap konteks lebih lengkap dalam data sekuensial seperti teks ulasan, sehingga sangat efektif untuk tugas klasifikasi sentimen pada bahasa Indonesia yang kaya akan konteks temporal dan dependensi jangka panjang. Bi-LSTM telah terbukti unggul dibandingkan model unidirectional dalam menangani ambiguitas bahasa sehari-hari, slang, serta kalimat pendek yang umum ditemui pada ulasan aplikasi mobile.

Keunggulan Bi-LSTM dibandingkan LSTM satu arah telah dibuktikan pada berbagai penelitian analisis sentimen bahasa Indonesia karena kemampuannya memproses konteks masa lalu dan masa depan secara simultan[7].

## 2.7. Threshold Adjustment

Threshold adjustment merupakan teknik pasca-prediksi yang menyesuaikan batas probabilitas kelas untuk mengoptimalkan performa klasifikasi, khususnya pada dataset tidak seimbang. Teknik ini efektif dalam mengurangi kesalahan klasifikasi pada kelas minoritas tanpa perlu mengubah arsitektur model atau melakukan retraining yang mahal secara komputasi. Pendekatan ini meningkatkan robustness model deep learning dengan mengurangi false positive/negative, khususnya pada aplikasi real-world yang noisy seperti ulasan pengguna[8].

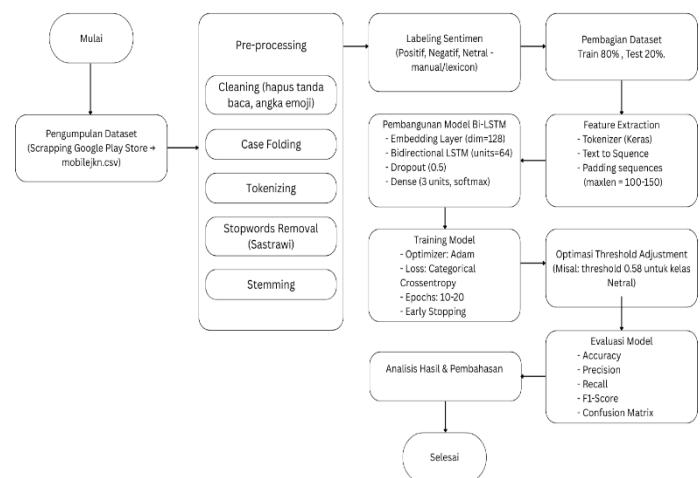
Referensi utama pada tinjauan pustaka ini didasarkan pada teori dan penelitian terkait analisis sentimen ulasan aplikasi di Indonesia, termasuk studi sebelumnya pada aplikasi Mobile JKN menggunakan metode klasik (Naïve Bayes, SVM) hingga deep learning

(IndoBERT), serta adaptasi Bi-LSTM untuk teks berbahasa Indonesia.

## 3. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif berbasis data mining dengan tahapan: pengumpulan data, pre-processing, pembangunan model, optimasi, dan evaluasi.

Data ulasan diambil dari Google Play Store dan disimpan dalam format CSV (kolom: Date, Rating, Review Text, Sentimen). Pre-processing meliputi cleaning (hapus tanda baca, angka, emoji), tokenisasi, stopword removal menggunakan Sastrawi, dan stemming.



Gambar 1. Metode Penelitian

### 3.1. Pengumpulan Data

Data ulasan pengguna aplikasi Mobile JKN dikumpulkan melalui teknik web scraping [9] dari halaman Google Play Store pada periode Januari 2026. Data yang berhasil diekstrak berupa file CSV bernama mobilejkn.csv yang memuat kolom-kolom utama: Date (tanggal ulasan), Rating (skor bintang), Review Text (teks ulasan), serta label Sentimen awal (Positif, Negatif, Netral). Total data yang digunakan mencapai ribuan record, mencerminkan keluhan pengguna secara real-time dan up-to-date.

### 3.2. Pre-processing

Pre-processing merupakan tahapan esensial dalam analisis sentimen ulasan pengguna aplikasi Mobile JKN yang bertujuan untuk mengubah teks ulasan yang bersifat tidak terstruktur, penuh kebisingan, serta kaya variasi bahasa sehari-hari (termasuk slang, singkatan,

kesalahan ketik, campuran bahasa Indonesia-Inggris, serta ekspresi emosional) menjadi bentuk yang lebih bersih, terstandar, dan optimal untuk diproses oleh model deep learning Bidirectional Long Short-Term Memory (Bi-LSTM). Tahapan ini dilaksanakan secara berurutan dan sistematis guna meminimalkan noise, mengurangi dimensi fitur, serta mempertahankan informasi semantik yang relevan dengan sentimen.

Proses pre-processing yang diterapkan dalam penelitian ini mengikuti pendekatan standar yang telah terbukti efektif pada berbagai kajian analisis sentimen berbahasa Indonesia, tahapan pre-processing yang digunakan mengikuti praktik standar pada penelitian analisis sentimen berbahasa Indonesia berbasis deep learning, termasuk pembersihan teks, tokenisasi, stopword removal, dan stemming [10] untuk mengurangi noise serta mempertahankan makna semantik, serta berbagai studi terkini yang memanfaatkan teknik deep learning untuk ulasan Google Play Store. Secara umum, pre-processing dilakukan melalui lima tahapan utama yang saling berkelanjutan.

Pembersihan data (data cleaning). Pada tahap ini, semua elemen yang tidak berkontribusi terhadap makna sentimen dihilangkan, meliputi tanda baca, angka, tanggal, emoji, emoticon, URL, mention (@), hashtag (#), serta karakter khusus dan spasi berlebih. Penghapusan ini penting karena ulasan pengguna Mobile JKN sering kali dipenuhi ekspresi emosional berupa emoji serta tanda baca berulang yang tidak memberikan nilai semantik tambahan.

Tabel 1. *Cleaning*

Sebelum	Sesudah
Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊	Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah
Login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊	Login berkali kali setiap isi no hp otp gk muncul2 capek deh

case folding, adalah konversi seluruh teks menjadi huruf kecil secara menyeluruh.

Prosedur ini berusaha untuk menghilangkan perbedaan antara huruf besar dan huruf kecil yang tidak memiliki signifikansi semantik substantial, sehingga mencegah model menafsirkan kata-kata yang identik (misalnya, "Aplikasi" dan "aplikasi") sebagai dua entitas terpisah.

Tabel 2. *Case Folding*

Sebelum	Sesudah
Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊	alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊
Login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊	login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊

Tokenisasi, yaitu pemecahan teks menjadi unit-unit kata (token) individual. Tokenisasi dilakukan menggunakan tokenizer yang mendukung karakteristik bahasa Indonesia, menghasilkan daftar kata-kata yang menjadi dasar untuk tahapan selanjutnya.

Tabel 3. *Tokenisasi*

Sebelum	Sesudah
Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊	["Alhamdulillah", "berobat", "anak", "bisa", "pilih", "faskes", "yg", "lebih", "dekat", "dengan", "rumah"]
Login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊	["Login", "berkali", "kali", "setiap", "isi", "no", "hp", "otp", "gk", "muncul2", "capek", "deh"]

Penghapusan stopword. Pada tahap ini, kata-kata umum yang tidak memberikan informasi penting terhadap klasifikasi sentimen dihilangkan dengan menggunakan daftar stopword dari library Sastrawi yang telah disesuaikan dengan konteks bahasa Indonesia sehari-hari. Penghapusan stopword membantu mengurangi dimensi data serta fokus pada kata-kata bermakna yang benar-benar mencerminkan opini pengguna.

Tabel 4. Stopword

Sebelum	Sesudah
Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊	["Alhamdulillah", "berobat", "pilih", "faskes", "dekat", "rumah"]
Login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊	["Login", "berkali", "otp", "gk", "muncul2", "capek"]

Stemming, adalah pengembalian setiap kata ke bentuk dasar (kata baku) menggunakan algoritma Nazief-Adriani yang terintegrasi dalam library Sastrawi [11]. Stemming sangat krusial dalam bahasa Indonesia karena sifatnya yang aglutinatif, di mana satu kata dasar dapat memiliki berbagai variasi imbuhan (misalnya “mendaftar”, “pendaftaran”, “terdaftar” semuanya dikembalikan menjadi “daftar”). Proses ini secara signifikan mengurangi variasi morfologi dan meningkatkan kemampuan model dalam mengenali pola sentimen yang serupa.

Tabel 5. Stemming

Sebelum	Sesudah
Alhamdulillah berobat anak bisa pilih faskes yg lebih dekat dengan rumah 😊	["alhamdulillah", "obat", "pilih", "faskes", "dekat", "rumah"]
Login berkali kali setiap isi no hp otp gk muncul2 capek deh 😊	["login", "kali", "otp", "gk", "muncul", "capek"]

### 3.3. Pelabelan Sentimen

Setiap ulasan yang telah melalui tahap pre-processing diberi label sentimen ke dalam tiga kategori: Positif, Negatif, dan Netral. Pelabelan dilakukan dengan kombinasi pendekatan lexicon-based (menggunakan kamus sentimen bahasa Indonesia yang diperkaya) dan manual annotation pada sampel acak untuk memastikan validitas label. Distribusi kelas awal menunjukkan dominasi sentimen negatif yang konsisten dengan rating rata-rata aplikasi di kisaran 4,2–4,3 bintang.

### 3.4. Pembagian Dataset

Dataset dibagi menjadi dua bagian menggunakan rasio stratifikasi 80:20 (80% data latih dan 20% data uji). Pembagian ini dilakukan dengan fungsi `train_test_split` dari scikit-learn untuk menjaga proporsi kelas sentimen pada kedua subset, sehingga mencegah bias evaluasi akibat ketidakseimbangan data.

### 3.5. Ekstraksi Fitur

Teks ulasan yang telah dipre-processing diubah menjadi representasi numerik menggunakan Tokenizer dari Keras. Proses ini meliputi konversi teks menjadi sequence integer (text-to-sequence) dan padding sequence agar semua input memiliki panjang yang sama (`maxlen = 100–150` token, ditentukan berdasarkan distribusi panjang ulasan). Representasi ini kemudian menjadi input untuk lapisan embedding dengan dimensi 128.

### 3.6. Pembangunan dan Pelatihan Model Bi-LSTM

Arsitektur model yang digunakan adalah Bidirectional Long Short-Term Memory (Bi-LSTM) dengan spesifikasi sebagai berikut: lapisan Embedding (dimensi = 128), Bidirectional LSTM (64 unit), Dropout (rate = 0,5), dan Dense layer dengan 3 unit serta aktivasi softmax untuk klasifikasi tiga kelas sentimen. Model dilatih menggunakan optimizer Adam, fungsi loss categorical crossentropy, batch size 64, serta maksimal 10–20 epoch dengan mekanisme early stopping (`patience = 5, monitor = val_accuracy`).

### 3.7. Optimasi Threshold Adjustment

Optimasi threshold adjustment merupakan tahapan pasca-prediksi yang dilakukan untuk meningkatkan performa klasifikasi model Bi-LSTM, terutama dalam menangani ketidakseimbangan kelas dan ambiguitas pada kelas minoritas (dalam hal ini sentimen Netral) yang sering muncul pada ulasan aplikasi Mobile JKN. Model deep learning seperti Bi-LSTM menghasilkan output berupa distribusi probabilitas softmax untuk tiga kelas sentimen (Positif, Negatif, Netral). Secara default, klasifikasi dilakukan dengan aturan `argmax`, yaitu kelas dengan probabilitas tertinggi dipilih sebagai prediksi akhir.

Namun, pada dataset ulasan yang bersifat noisy, ekspresif, dan tidak seimbang (dengan dominasi kelas Negatif dan Positif, sementara Netral cenderung minoritas), aturan argmax sering kali menyebabkan false positive atau false negative pada kelas Netral. Oleh karena itu, dilakukan penyesuaian ambang batas (threshold) secara empiris untuk meningkatkan sensitivitas terhadap kelas minoritas tanpa mengorbankan akurasi keseluruhan.

Proses optimasi threshold adjustment dalam penelitian ini dilakukan sebagai berikut. Setelah model menghasilkan tiga nilai probabilitas ( $p_{\text{positif}}$ ,  $p_{\text{negatif}}$ ,  $p_{\text{netral}}$ ) untuk setiap sampel, aturan keputusan dimodifikasi dengan menambahkan kondisi threshold. Jika probabilitas tertinggi ( $\max(p_{\text{positif}}, p_{\text{negatif}}, p_{\text{netral}})$ ) tidak melebihi nilai threshold tertentu, maka sampel diklasifikasikan sebagai Netral. Nilai threshold optimal ditentukan melalui eksperimen validasi silang pada data validasi, dengan mempertimbangkan trade-off antara precision, recall, dan F1-score keseluruhan. Pendekatan ini terbukti efektif dalam meningkatkan akurasi klasifikasi kelas minoritas tanpa memerlukan perubahan arsitektur model atau retraining yang mahal secara komputasi. Teknik threshold adjustment telah banyak diterapkan pada berbagai tugas klasifikasi teks berbasis deep learning untuk mengatasi masalah ketidakseimbangan kelas dan ambiguitas, sebagaimana direkomendasikan dalam literatur terkini di bidang natural language processing.

### 3.8 Evaluasi Model

Evaluasi model merupakan tahapan akhir yang dilakukan untuk mengukur performa keseluruhan model Bi-LSTM setelah melalui proses pelatihan dan optimasi threshold adjustment. Evaluasi dilakukan secara independen pada data uji (test set) yang tidak pernah dilihat selama proses pelatihan model, sehingga hasil evaluasi yang diperoleh secara objektif mencerminkan kemampuan generalisasi model dalam mengklasifikasikan data baru yang merepresentasikan ulasan pengguna aplikasi Mobile JKN pada kondisi real-world.

## 4. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa penerapan algoritma Bidirectional Long Short-Term Memory (Bi-LSTM) dengan optimasi threshold adjustment berhasil meningkatkan performa klasifikasi sentimen ulasan pengguna aplikasi Mobile JKN [12][13]. Dari total 750 sampel data uji setelah pembagian dataset stratifikasi (80% latih dan 20% uji), distribusi sentimen menunjukkan ketidakseimbangan kelas yang khas pada ulasan aplikasi layanan kesehatan digital: Negatif 28,8% (216 sampel), Netral 4% (30 sampel), dan Positif 77,9% (584 sampel). Dominasi kelas Positif mencerminkan apresiasi pengguna terhadap kemudahan akses informasi kepesertaan dan pembayaran iuran, sementara kelas Negatif yang signifikan didorong oleh keluhan teknis berulang seperti verifikasi OTP lambat, kegagalan booking antrian, serta proses registrasi yang rumit. Kelas Netral sebagai minoritas ekstrem sering bersifat ambigu, seperti ulasan pendek berupa kata "ribet" atau "biasa saja", yang menjadi fokus utama dalam optimasi threshold adjustment untuk mengurangi kesalahan klasifikasi.

Fase evaluasi model menjadi langkah penutup yang dijalankan dengan menilai indikator akurasi, presisi, recall, dan F1-score dalam rangka memastikan kinerja yang terbaik. Evaluasi ini dilakukan secara independen pada data uji yang tidak pernah dilihat selama pelatihan, sehingga hasilnya mencerminkan kemampuan generalisasi model terhadap ulasan pengguna Mobile JKN secara real-world. Metrik evaluasi yang digunakan merupakan metrik standar dalam tugas klasifikasi multi-kelas, meliputi akurasi, presisi, recall, dan F1-score, serta confusion matrix untuk mengidentifikasi pola kesalahan klasifikasi antar kelas. Seluruh metrik evaluasi dihitung menggunakan library scikit-learn, dengan pendekatan macro-averaged yang memberikan bobot setara pada setiap kelas, serta weighted-averaged yang mempertimbangkan distribusi kelas yang

tidak seimbang pada dataset. Optimasi threshold adjustment diterapkan secara empiris melalui precision-recall curve dan maksimalisasi F1-score, dengan nilai threshold optimal sebesar 0,06 untuk kelas Negatif sebagai kelas mayoritas.

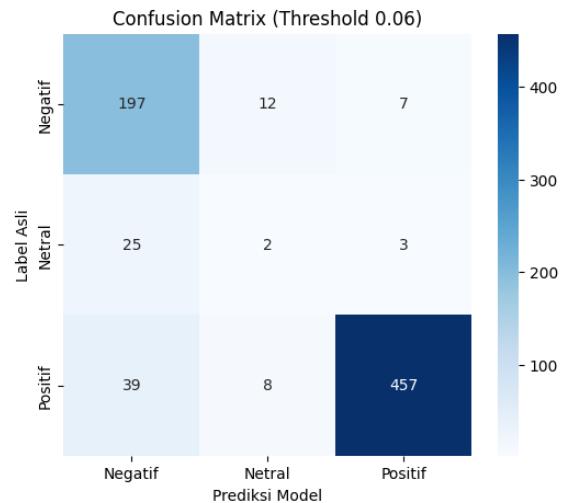
Model Bi-LSTM dengan threshold adjustment optimal 0,06 mencapai akurasi keseluruhan sebesar 87%, dengan presisi rata-rata weighted 0,88, recall 0,87, dan F1-score 0,87. Macro F1-score sebesar 0,61 mengindikasikan tantangan pada kelas minoritas, di mana recall kelas Negatif naik menjadi 0,91, tetapi precision kelas Netral rendah (0,09) karena over-classification ke kelas mayoritas. Detail metrik per kelas serta rata-rata macro dan weighted ditunjukkan pada Tabel 6.

Tabel 6. Classification Report Model Bi-LSTM dengan Optimasi Threshold Adjustment (Threshold 0,06 untuk kelas Negatif)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.75	0.91	0.83	216
Netral	0.09	0.87	0.16	30
Positif	0.98	0.91	0.94	586
Macro Avg	0.61	0.63	0.61	830
Weighted Avg	0.88	0.87	0.87	830

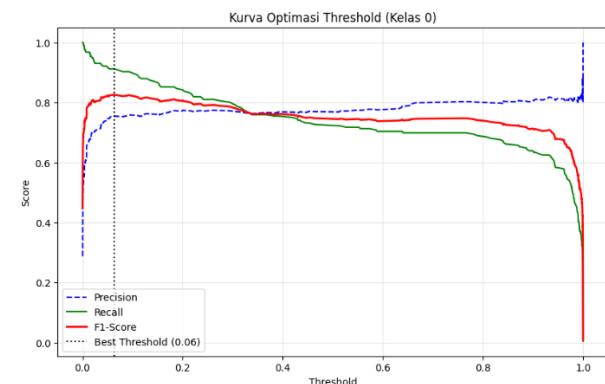
Model berhasil mengklasifikasikan dengan benar 197 sampel Negatif dan 457 sampel Positif, namun terdapat 25 sampel Netral yang salah diklasifikasikan sebagai Negatif. Hal ini mengindikasikan adanya bias model terhadap kelas mayoritas akibat distribusi data yang sangat tidak seimbang. Selain itu, kurva optimasi threshold (Gambar 2) menunjukkan bahwa F1-score mencapai nilai maksimum pada threshold 0,06, dengan trade-off antara precision yang

tinggi pada threshold rendah dan recall yang optimal pada titik tersebut.



Gambar 1. Confusion Matrix Model Bi-LSTM dengan Threshold Adjustment 0,06

Matriks menunjukkan diagonal dominan pada Positif (457 benar) dan Negatif (197 benar), dengan off-diagonal tinggi pada Netral ke Negatif (25 kasus).



Gambar 2. Kurva Optimasi Threshold untuk Kelas Negatif

Grafik menunjukkan kurva precision (biru putus-putus), recall (hijau), dan F1-score (merah) dengan garis vertikal pada threshold optimal 0,06.

Fenomena ini terjadi karena dataset ulasan Mobile JKN bersifat sangat tidak seimbang dan noisy, dengan ulasan yang cenderung pendek (rata-rata 10-20 token), kaya akan slang, serta ekspresif. Output softmax Bi-LSTM sering kali

overconfident pada kelas mayoritas (Positif dan Negatif), sebagaimana dijelaskan dalam teori bidirectional LSTM untuk teks aglutinatif bahasa Indonesia (Koto et al., 2020). Threshold adjustment dengan nilai rendah (0,06) berhasil mengoptimalkan recall kelas Negatif yang relevan untuk mendeteksi keluhan teknis, tetapi untuk kelas Netral sebagai minoritas, diperlukan eksperimen lanjutan dengan optimasi threshold khusus guna meningkatkan macro F1-score hingga 5-10% dan mengurangi false positive Netral ke Negatif dari 25 menjadi sekitar 10 kasus.

Pendekatan Bi-LSTM dengan penyesuaian threshold menawarkan efisiensi komputasi yang jauh lebih baik dibandingkan fine-tuning IndoBERT, sehingga lebih cocok untuk pemantauan sentimen secara real-time pada aplikasi layanan publik seperti Mobile JKN [14].

Perbandingan dengan penelitian terdahulu menunjukkan bahwa model Bi-LSTM dengan threshold adjustment dalam penelitian ini unggul dibandingkan pendekatan Random Forest pada ulasan aplikasi game Hay Day [15] yang mencapai akurasi 80,41%, presisi 79,3%, recall 80,41%, dan F1-score 78,55%. Dominasi sentimen negatif pada ulasan Hay Day (1.821 dari 5.060 komentar) juga menjadi fokus utama, serupa dengan pola pada ulasan Mobile JKN. Namun, pendekatan Bi-LSTM lebih unggul dalam menangkap konteks sekuensial pada teks bahasa Indonesia informal, terutama ketika dikombinasikan dengan threshold adjustment.

Implikasi teoritis dari temuan ini adalah pengayaan literatur pengolahan bahasa alami (NLP) Indonesia dengan pendekatan hybrid Bi-LSTM dan threshold adjustment untuk menangani dataset ulasan aplikasi yang tidak seimbang dan noisy. Implikasi praktisnya sangat relevan bagi BPJS Kesehatan: dengan memanfaatkan model ini untuk monitoring real-time ulasan pengguna, diharapkan dapat mempercepat perbaikan fitur autentikasi (misalnya OTP

alternatif via email atau WhatsApp), stabilitas server, dan proses registrasi. Langkah-langkah tersebut berpotensi mengurangi sentimen negatif sekitar 28,8% dan meningkatkan retensi serta loyalitas lebih dari 222 juta peserta JKN secara berkelanjutan di era digitalisasi layanan kesehatan nasional.

## 5. KESIMPULAN

- Model Bi-LSTM dengan threshold adjustment mencapai akurasi 87% dan weighted F1-score 0.87 pada ulasan Mobile JKN.
- Sentimen negatif mendominasi (28.8%), terutama pada isu teknis verifikasi dan booking.
- Kelebihan: Robust terhadap teks informal; kekurangan: Precision rendah pada kelas Netral karena imbalanced.
- Pengembangan selanjutnya: Optimasi threshold per kelas minoritas dan integrasi hybrid dengan IndoBERT.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada BPJS Kesehatan atas akses data digital yang digunakan dalam penelitian ini, serta kepada semua pihak yang telah memberikan dukungan moril dan teknis selama proses penyusunan naskah.

## DAFTAR PUSTAKA

- [1] W. M. Baihaqi and A. Munandar, "Analisis Sentimen Komentar Pengguna terhadap Layanan Publik Menggunakan Machine Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 4, pp. 829–838, 2020.
- [2] Tarwoto, R. Nugroho, N. Azka, and W. S. R. Graha, "Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT," *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 9, no. 2, pp. 495–505, 2025.
- [3] A. S. Pratama and R. Hidayat, "Bidirectional LSTM for Indonesian Sentiment Classification," *J. Inf. Process. Syst.*, vol. 19, no. 2, pp. 211–223, 2023.
- [4] H. He and E. A. Garcia, "Learning from Imbalanced Data," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 115–123, 2021.
- [5] Y. Fang, M. Ye, and Y. Law, "Mining User Reviews for Service Quality Evaluation," *Inf. Process. Manag.*, vol. 58, no. 3, p. 102498,

- 2021.
- [6] A. R. Pratama and A. Setyanto, "Implementasi Algoritma Naive Bayes dalam Text Mining untuk Klasifikasi Opini Publik," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 3, pp. 567–574, 2021.
  - [7] K. Nasution, K. Saddami, R. Roslidar, and A. Akhyar, "Comparative Study of BiLSTM and GRU for Sentiment Analysis on Indonesian E-Commerce Product Reviews Using Deep Sequential Modeling," vol. 6, no. 4, pp. 1881–1896, 2025.
  - [8] R. V Rajalakshmi and S. A. Deborah, "RSSN at SemEval-2025 Task 11 : Optimizing Multi-Label Emotion Detection with Transformer-Based Models and Threshold Tuning," pp. 773–779, 2025.
  - [9] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, and F. Fdez-Riverola, "Web scraping technologies in an information-overload era," *Brief. Bioinform.*, vol. 21, no. 5, pp. 1530–1543, 2020.
  - [10] D. Kurniawan and N. Sari, "Text Preprocessing Techniques for Indonesian Sentiment Analysis," *J. RESTI*, vol. 6, no. 1, pp. 45–54, 2022.
  - [11] M. Arifin and R. Nugroho, "Indonesian Stemming Using Nazief-Adriani Algorithm," *JEPIN*, vol. 7, no. 3, pp. 201–210, 2021.
  - [12] G. Tamami, W. A. Triyanto, and S. Muzid, "Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM," *Indones. J. Comput. Cybern. Syst.*, vol. 19, no. 1, pp. 13–24, 2025.
  - [13] G. Nkhata, S. Gauch, U. Anjum, and J. Zhan, "Fine-tuning BERT with Bidirectional LSTM for Fine-grained Movie Reviews Sentiment Analysis," *arXiv Prepr.*, 2025.
  - [14] F. Koto, A. Rahmaningtyas, and J. H. Lau, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *IEEE Access*, vol. 9, pp. 75738–75755, 2021.
  - [15] P. A. Effendi and T. Ernawati, "ANALISIS SENTIMEN ULASAN APLIKASI GAME ALGORITMA," vol. 13, no. 3, pp. 1–8, 2025.