

Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi JConnect Mobile Menggunakan IndoBERT dan mBERT

Aspect-Based Sentiment Analysis of JConnect Mobile Application Reviews Using IndoBERT and mBERT

Nadin Isna Monica^{*1}, Abdul Rezha Efrat Najaf², Nambi Sembilu³ (* correspondent author)

^{1,2,3}Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur

E-mail: ¹nadinimo.desk@gmail.com, ²rezha.efrat.sifo@upnjatim.ac.id,

³nambi.si@upnjatim.ac.id

Abstrak

JConnect Mobile merupakan aplikasi mobile banking milik Bank Jatim yang menyediakan berbagai layanan transaksi perbankan secara digital. Meskipun jumlah pengguna terus meningkat, masih terdapat berbagai keluhan terkait tampilan, fitur dan performa, serta layanan yang diberikan, yang tercermin dari ulasan pengguna dan rating aplikasi yang relatif rendah. Penelitian ini bertujuan untuk membandingkan performa model IndoBERT dan mBERT dalam klasifikasi sentimen berbasis aspek pada ulasan pengguna JConnect Mobile. Data penelitian yang digunakan sebanyak 4.344 ulasan pengguna periode 2021–2025 yang diperoleh dari Google Play Store dan App Store. Tahapan penelitian meliputi data collecting, labeling, preprocessing, feature extraction, modeling, dan evaluasi. Eksperimen dilakukan menggunakan metode 5-Fold Cross Validation dengan 24 skenario pengujian yang terdiri atas konfigurasi hyperparameter epoch 5 dan 10, dropout 0,2, 0,4, dan 0,6, serta batch size 16 dan 32. Hasil evaluasi 5-Fold Cross Validation menunjukkan bahwa IndoBERT memberikan performa terbaik dengan mean F1-score sebesar 0,86 pada aspek tampilan, 0,83 pada aspek fitur dan performa, serta 0,74 pada aspek layanan. Berdasarkan hasil tersebut, IndoBERT dipilih sebagai model terbaik karena konsisten mengungguli mBERT pada seluruh aspek. Evaluasi pada data uji menunjukkan bahwa model terbaik mampu melakukan klasifikasi sentimen dengan baik dengan F1-score sebesar 0,80 pada aspek tampilan, 0,83 pada aspek fitur dan performa, serta 0,81 pada aspek layanan.

Kata kunci: JConnect Mobile, Aspect-Based Sentiment Analysis (ABSA), IndoBERT, mBERT, K-Fold Cross Validation

Abstract

JConnect Mobile is a mobile banking application developed by Bank Jatim that provides various digital banking transaction services. Although the number of users has continued to increase, the application still receives numerous complaints regarding its interface, features and performance, and service quality, as reflected in user reviews and its relatively low application rating. This study aims to compare the performance of the IndoBERT and mBERT models for aspect-based sentiment classification of JConnect Mobile user reviews. The dataset consists of 4,344 user reviews collected from Google Play Store and App Store between 2021 and 2025. The research methodology includes data collection, data labeling, preprocessing, feature extraction, modeling, and evaluation. The experiments were conducted using 5-fold cross-validation with 24 experimental scenarios involving combinations of hyperparameters,

including 5 and 10 epochs, dropout rates of 0.2, 0.4, and 0.6, and batch sizes of 16 and 32. The results of the 5-fold cross-validation show that IndoBERT achieved the best performance, with mean F1-scores of 0.86 for the interface aspect, 0.83 for the features and performance aspect, and 0.74 for the service aspect. Based on these results, IndoBERT was selected as the best-performing model because it consistently outperformed mBERT across all aspects. Evaluation on the test dataset further demonstrated that the best-performing model achieved F1-scores of 0.80 for the interface aspect, 0.83 for the features and performance aspect, and 0.81 for the service aspect.

Keywords: JConnect Mobile, Aspect-Based Sentiment Analysis (ABSA), IndoBERT, mBERT, K-Fold Cross Validation

1. PENDAHULUAN

Peningkatan ekspektasi nasabah terhadap kemudahan, kecepatan, dan keamanan layanan perbankan telah mendorong transformasi digital menjadi suatu keharusan strategis bagi Bank Pembangunan Daerah (BPD) [1]. Pada tahun 2021, Bank Jatim meluncurkan aplikasi digital JConnect Mobile sebagai bagian dari transformasi layanan perbankan menuju era digital. Aplikasi ini memungkinkan nasabah melakukan berbagai transaksi perbankan secara daring dengan lebih mudah dan praktis. [2]. Aplikasi JConnect Mobile telah diunduh oleh lebih dari satu juta pengguna melalui Google Play Store. Meskipun berbagai fitur dan layanan yang tersedia mendapat apresiasi dari pengguna, sejumlah keluhan masih ditemukan pada ulasan aplikasi, seperti keterlambatan transaksi, kesulitan menghubungi layanan pelanggan, serta permasalahan pada tampilan antarmuka [3]. Meskipun adopsi JConnect Mobile terus meningkat, ulasan pengguna menunjukkan bahwa kualitas layanan belum sepenuhnya memenuhi harapan. Kondisi ini tercermin dari rating aplikasi sebesar 3,0 berdasarkan 17,6 ribu ulasan di Google Play Store per November 2025, yang masih relatif rendah dibandingkan aplikasi digital banking BPD lainnya [4].

Urgensi penelitian sejalan dengan arah strategis Bank Jatim dalam Rencana Strategis Teknologi Informasi (RSTI) 2024–2028, khususnya pada fase tahun 2027 yakni *Customer Empowerment through AI and ML* yang menekankan pengalaman pengguna dan pengambilan keputusan berbasis data [5]. Maka diperlukan pemanfaatan teknologi kecerdasan buatan yang mampu mengolah ulasan pengguna secara efektif. Salah satu pendekatan yang dapat digunakan adalah *Aspect-Based Sentiment Analysis* (ABSA), yang memungkinkan identifikasi sentimen pengguna secara lebih spesifik berdasarkan aspek-aspek tertentu dari layanan digital. Dengan demikian, informasi yang diperoleh tidak hanya menunjukkan sentimen secara umum, tetapi lebih spesifik terhadap aspek yang dianalisis. Penelitian ini mendukung upaya Bank Jatim dalam memahami kebutuhan dan persepsi nasabah secara lebih mendalam sehingga dapat menjadi dasar dalam evaluasi dan peningkatan kualitas layanan digital yang berorientasi pada kebutuhan pelanggan.

Penerapan ABSA telah banyak diterapkan pada aplikasi perbankan. Penelitian pada ulasan aplikasi Flip menggunakan metode *Support Vector Machine*

(SVM) menunjukkan bahwa ABSA dapat mengidentifikasi sentimen pengguna pada setiap aspek yang telah ditentukan dengan memperoleh hasil tingkat akurasi yang konsisten di atas 89% [6]. Selain itu, penelitian analisis sentimen perbankan menunjukkan keunggulan model *transfer learning* pada analisis sentimen pada aplikasi Bank Syariah Indonesia (BSI) yang membandingkan SVM dan IndoBERT yang menunjukkan bahwa IndoBERT memiliki kinerja lebih baik dengan akurasi 85% [7]. Penelitian lain pada ulasan aplikasi BRImo menggunakan IndoBERT juga menunjukkan performa klasifikasi sentimen yang tinggi mencapai akurasi 95% yang menunjukkan model menghasilkan prediksi yang akurat [8].

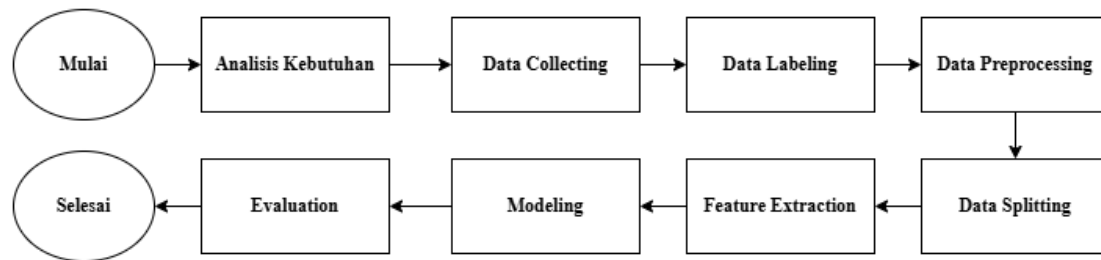
Beberapa penelitian sebelumnya telah mencoba melakukan analisis sentimen terhadap ulasan aplikasi JConnect Mobile dengan pendekatan Machine learning klasik. Penelitian oleh [9] berfokus pada analisis sentimen berdasarkan aspek menggunakan algoritma *Naïve Bayes*. Penelitian oleh [10] berfokus pada mengevaluasi performa algoritma *Multinomial Naive Bayes* dan *Support Vectore Machine* dengan memperoleh performa terbaik pada algoritma SVM.

Penelitian terdahulu pada ulasan JConnect Mobile masih didominasi oleh metode *machine learning* klasik yang memiliki keterbatasan dalam memahami konteks dan hubungan semantik antar kata [11]. Oleh karena itu, model transfer learning seperti IndoBERT dan mBERT menjadi alternatif yang lebih relevan karena mampu menangkap konteks bahasa secara dua arah (bidirectional) serta telah dilatih pada korpus bahasa yang besar [12][13][14]. Selain itu, penelitian sebelumnya umumnya hanya memanfaatkan data Google Play Store, sedangkan penelitian ini menggunakan ulasan dari Google Play Store dan App Store periode 2021–2025 untuk memperoleh representasi opini pengguna yang lebih representatif.

Berbeda dengan penelitian sebelumnya, penelitian ini menggunakan kombinasi data ulasan pengguna JConnect Mobile yang diperoleh dari Google Play Store dan App Store periode 2021–2025 serta membandingkan performa IndoBERT dan mBERT untuk analisis sentimen berbasis aspek menggunakan *hyperparameter tuning* dan *5-Fold Cross Validation*. Pendekatan ini menjadi kebaruan penelitian dengan sumber data yang lebih beragam dan pendekatan evaluasi yang komprehensif dan *robust* untuk memperoleh konfigurasi model terbaik.

2. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan pada penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

2.1. Analisis Kebutuhan

Analisis kebutuhan merupakan tahap yang dilakukan untuk mengidentifikasi kebutuhan data, *hardware*, dan *software* yang diperlukan untuk mendukung proses penelitian. Kebutuhan data pada penelitian ini berupa ulasan pengguna aplikasi JConnect Mobile yang diperoleh dari Google Play Store dan App Store. Data yang digunakan merupakan ulasan yang dipublikasikan pada periode 17 Agustus 2021 hingga 31 Desember 2025. *Hardware* yang digunakan dalam penelitian ini adalah sebuah laptop Acer Aspire A514-54G dengan spesifikasi CPU Intel Core i5-1135G7, RAM 8 GB, SSD 512 GB, dan sistem operasi Windows 11 Pro 64-bit. Proses pelatihan (*fine-tuning*) dilakukan menggunakan Kaggle Notebook dengan akselerasi GPU Tesla T4 (CUDA). *Software* yang digunakan meliputi bahasa pemrograman *Python*, *framework PyTorch* versi 2.10.0+cu128, *library Transformers* versi 4.41.2, *Accelerate* versi 0.30.1, *PEFT* versi 0.11.1, serta *library* pendukung seperti *Scikit-learn*, *Pandas*, dan *NumPy*.

2.2. Data Collecting

Data ulasan JConnect Mobile dikumpulkan menggunakan metode *scraping* dari *Google Play Store* dan *App Store* menggunakan *library Python google-play-scraper* dan *requests*. Pengambilan data dilakukan pada periode 17 Agustus 2021 hingga 31 Desember 2025. Selanjutnya, dilakukan proses *filtering* untuk meningkatkan kualitas dataset dengan menghapus ulasan yang hanya berisi emoji, ulasan duplikat, dan ulasan yang tidak informatif atau tidak relevan. Proses ini bertujuan untuk memastikan data yang digunakan pada tahap analisis memiliki kualitas yang baik dan representatif.

2.3. Data Labelling

Pelabelan data dilakukan secara manual oleh tiga anotator berdasarkan aspek dan sentimen yang terkandung dalam ulasan. Tingkat kesepakatan antar anotator diukur menggunakan *Krippendorff's Alpha* untuk memastikan reliabilitas pelabelan. Setelah nilai *alpha* memenuhi batas minimal ($\alpha \geq 0,67$). Rumus dasar *Krippendorff's Alpha* sebagai berikut [15].

$$\alpha = 1 - \frac{D_o}{D_e} \quad (1)$$

Keterangan:

α : Nilai *Krippendorff's Alpha*

D_o : *Observed Disagreement*, ketidaksepakatan aktual antar penilai

D_e : *Expected Disagreement*, ketidaksepakatan yang diharapkan

Label akhir ditentukan menggunakan metode *majority voting*, yaitu label yang disepakati oleh minimal dua dari tiga anotator. Pada penelitian ini digunakan tiga aspek, yaitu tampilan, fitur dan performa, serta layanan. Setiap aspek kemudian diklasifikasikan ke dalam empat kelas sentimen, yaitu positif, negatif, netral, dan none, di mana kelas none menunjukkan bahwa aspek tersebut tidak dibahas dalam ulasan.

2.4. Data Preprocessing

Tahap *preprocessing* bertujuan untuk membersihkan serta menyiapkan data teks agar dapat diolah oleh model secara optimal. Tahapan yang diterapkan meliputi *case folding*, *cleaning*, dan normalisasi.

a. Case Folding

Case folding merupakan proses mengubah seluruh huruf pada teks menjadi huruf kecil (*lowercase*) untuk menyeragamkan format penulisan dan mengurangi variasi penggunaan huruf.

b. Cleaning

Tahap *cleaning* merupakan proses menghapus elemen-elemen yang tidak relevan, seperti URL, hashtag, emoji, tanda baca, karakter nonalfabet, dan spasi berlebih, sehingga menghasilkan teks yang lebih bersih dan terstruktur.

c. Normalisasi

Tahap normalisasi merupakan proses mengubah kata tidak baku, *typo*, singkatan, *slang*, dan karakter berulang menjadi bentuk baku menggunakan kamus normalisasi agar representasi kata menjadi lebih konsisten.

2.5. Data Splitting

Dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 90:10 menggunakan metode *stratified sampling*. Metode ini digunakan untuk mempertahankan proporsi kelas pada setiap aspek sehingga distribusi data pada kedua subset tetap representatif terhadap dataset asli.

2.6. Feature Extraction

Tahap *feature extraction* bertujuan mentransformasikan teks hasil *preprocessing* menjadi representasi numerik yang dapat diproses oleh model. Pada penelitian ini, ekstraksi fitur menggunakan model IndoBERT dan mBERT yang menghasilkan *contextual embeddings* melalui arsitektur *Transformer*. Teks ulasan terlebih dahulu ditokenisasi menggunakan tokenizer berbasis *WordPiece*, kemudian setiap token direpresentasikan dalam bentuk vektor kontekstual. Representasi vektor yang dihasilkan selanjutnya digunakan sebagai input pada proses *fine-tuning* untuk klasifikasi sentimen berbasis aspek.

2.7. Modeling

Tahap modeling dilakukan menggunakan IndoBERT dan mBERT untuk membangun model klasifikasi sentimen berbasis aspek. Sebelum proses pelatihan, dilakukan penanganan *class imbalance* menggunakan *oversampling* digunakan pada

data latih dengan meningkatkan jumlah sampel pada kelas minoritas sehingga distribusi data menjadi lebih seimbang. Selanjutnya, pada proses pelatihan diterapkan *Focal Loss* berfungsi untuk memberikan bobot yang lebih besar terhadap sampel yang sulit diklasifikasikan. Penggunaan *oversampling* dan *Focal Loss* dilakukan pada tahapan yang berbeda namun saling melengkapi. Selain itu, dilakukan eksperimen *hyperparameter* untuk memperoleh konfigurasi terbaik pada masing-masing model. Evaluasi dilakukan menggunakan metode *5-fold cross validation* guna menghasilkan performa yang lebih stabil dan mengurangi bias akibat pembagian data tertentu.

Tabel 1. Skenario Eksperimen

No.	Model Klasifikasi	Kombinasi Parameter	Jumlah Skenario
1.	IndoBERT	Epoch = [5, 10] Batch Size = [16, 32] Dropout = [0.2, 0.4, 0.6]	12
2.	Multilingual BERT	Epoch = [5, 10] Batch Size = [16, 32] Dropout = [0.2, 0.4, 0.6]	12
Total			24

Setelah diperoleh konfigurasi optimal, dilakukan *final training*. Pada tahap ini diterapkan mekanisme *early stopping* diterapkan untuk mencegah terjadinya *overfitting* selama proses pelatihan model dan memperoleh model terbaik yang selanjutnya digunakan pada tahap pengujian data test.

2.8. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur kinerja model dalam melakukan klasifikasi sentimen. Berikut adalah persamaan dari metrik evaluasi.

- a. *Accuracy* merupakan metrik yang digunakan untuk menilai tingkat ketepatan model dengan membandingkan jumlah prediksi yang benar terhadap total data pada proses pengujian.

$$Accuracy = \frac{TPos+TNet+TNeg}{TPos+FPos+TNet+FNeg+TNeg+FNeg} \quad (2)$$

- b. *Precision* merupakan metrik yang mengukur proporsi prediksi positif yang benar terhadap seluruh data yang diprediksi sebagai kelas positif.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

- c. *Recall* merupakan metrik yang mengukur kemampuan model dalam mendeteksi data kelas positif dengan membandingkan jumlah prediksi positif yang benar terhadap seluruh data yang sebenarnya positif.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

- d. *F1-Score* merupakan metrik yang digunakan untuk mengevaluasi keseimbangan antara *precision* dan *recall*, yang dihitung menggunakan rata-rata harmonik dari kedua metrik tersebut.

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

3. HASIL DAN PEMBAHASAN

3.1. Data Collecting

Pengumpulan data ulasan pengguna JConnect Mobile diperoleh dari Google Play Store dan App Store. Hasil pengumpulan data diperoleh sebanyak 5.372 ulasan dari Google Play Store dan 296 ulasan dari App Store, sehingga total data yang berhasil dikumpulkan sebanyak 5.668 ulasan. Tahap berikutnya adalah *filtering*, yang bertujuan meningkatkan kualitas data agar layak digunakan pada proses pelabelan dan pemodelan. Tahap pertama adalah penghapusan ulasan yang hanya berisi emoji sehingga jumlah data berkurang dari 5.668 menjadi 5.637 ulasan. Tahap kedua adalah penghapusan data duplikat yang menghasilkan 4.970 ulasan. Tahap terakhir adalah penghapusan ulasan yang tidak informatif dan tidak relevan dengan tujuan penelitian. Setelah proses *filtering* selesai, jumlah data yang digunakan dalam penelitian ini sebanyak 4.344 ulasan. Berikut adalah data yang dihapus pada tahap *filtering*.

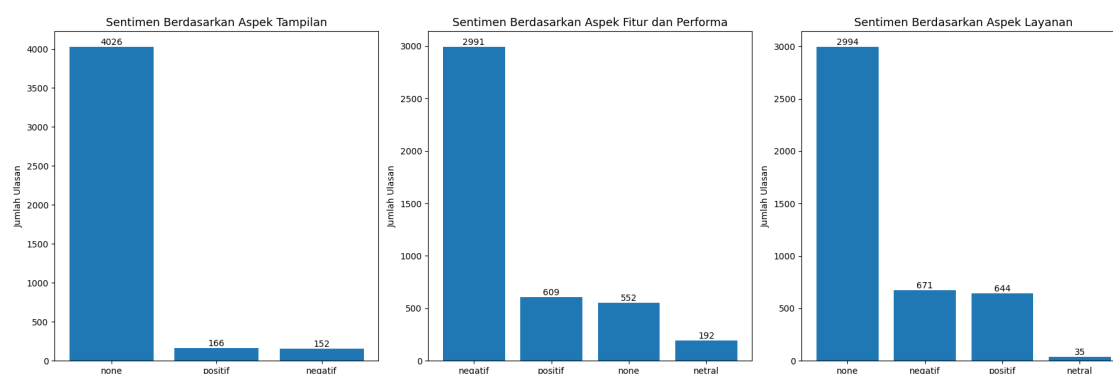
Tabel 2. Data Filtering

Filtering	Ulasan
Hanya Emoji	
Hanya Emoji	
Duplikat	sangat membantu
Duplikat	aplikasi sering eror
Tidak Relevan	Nana nasir rotun nadillah anggun Megasari khoiruululazmi ibu srihartatik bidan

Tidak Relevan	" rpbebasfinansialrp " untuk 25k Gratis di aplikasi Bibit. Bank jatim bank andalan arek jawa timur
---------------	--

3.2. Data Labelling

Pelabelan data dilakukan secara manual oleh tiga orang anotator dengan mengacu pada pedoman pelabelan yang telah disusun sebelumnya. Setiap ulasan diberikan label aspek (tampilan, fitur dan performa, serta layanan) dan label sentimen (positif, negatif, netral, atau none). Untuk mengukur tingkat kesepakatan antar anotator, digunakan metode *Krippendorff's Alpha*. Hasil pengujian menunjukkan nilai Krippendorff's Alpha sebesar 0,764 pada aspek Tampilan, 0,888 pada aspek Fitur dan Performa, serta 0,837 pada aspek Layanan. Berdasarkan kriteria *Krippendorff's Alpha*, seluruh aspek memperoleh nilai di atas 0,67 yang menunjukkan tingkat reliabilitas yang baik dan dapat diterima untuk penelitian. Setelah proses pelabelan selesai, dilakukan *majority voting* untuk menentukan label akhir pada setiap ulasan berdasarkan suara mayoritas dari ketiga anotator. Distribusi sentimen hasil pelabelan pada masing-masing aspek ditunjukkan pada Gambar 2.



Gambar 2. Distribusi Sentimen Per Aspek

Berdasarkan visualisasi distribusi sentimen pada masing-masing aspek pada Gambar 2, menunjukkan bahwa sebagian besar ulasan pengguna tidak mengandung sentimen (none), terutama pada aspek Tampilan dan Layanan. Pada aspek Fitur dan Performa, sentimen negatif mendominasi dibandingkan sentimen positif, yang menunjukkan bahwa pengguna masih banyak memberikan keluhan terkait fungsi dan kinerja aplikasi. Sementara itu, pada aspek Layanan, jumlah sentimen positif dan negatif relatif seimbang, meskipun sentimen negatif sedikit lebih tinggi. Distribusi ini menunjukkan adanya ketidakseimbangan kelas pada dataset, yang ditandai perbedaan jumlah data yang cukup signifikan antar kelas sentimen.

3.3. Data Preprocessing

Berikut adalah hasil dari proses data preprocessing meliputi *case folding*, *cleaning*, dan normalisasi.

Tabel 3. Data Preprocessing

Ulasan Asli	saya sangat terbantu dengan Jconnect Bank Jatim terutama untuk bayar PBB, bayar Pajak Kendaraan Bermotor, BPHTB, PDAM, PLN, Pulsa,Wifi, QRIS . Everyday gunain Jconnect untuk segala trx 👍👍👍👍👍
Case Folding	saya sangat terbantu dengan jconnect bank jatim terutama untuk bayar pbb, bayar pajak kendaraan bermotor, bphtb, pdam, pln, pulsa,wifi, qris . everyday gunain jconnect untuk segala trx 👍👍👍👍👍
Cleaning	saya sangat terbantu dengan jconnect bank jatim terutama untuk bayar pbb bayar pajak kendaraan bermotor bphtb pdam pln pulsa wifi qris everyday gunain jconnect untuk segala transaksi
Normalisasi	saya sangat terbantu dengan jconnect bank jatim terutama untuk bayar pbb bayar pajak kendaran bermotor bphtb pdam pln pulsa wifi qris everyday menggunakan jconnect untuk segala transaksi

3.4. Data Splitting

Dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 90:10 menggunakan teknik *stratified sampling*. Dari total 4.344 data, diperoleh 3.909 data latih dan 435 data uji. *Stratified sampling* digunakan untuk mempertahankan proporsi distribusi kelas pada setiap aspek sehingga data latih dan data uji tetap mencerminkan distribusi data asli.

3.5. Feature Extraction

Pada penelitian ini, tokenisasi dilakukan menggunakan *AutoTokenizer* dari pustaka *Transformers*. *AutoTokenizer* secara otomatis memuat *tokenizer* yang sesuai dengan model IndoBERT maupun mBERT, kemudian mengubah teks menjadi *token*, *input IDs*, dan *attention mask* sebagai masukan bagi model.

3.6. Modeling

Pelatihan model dilakukan menggunakan dua model *transfer learning*, yaitu IndoBERT dan mBERT. Untuk memperoleh konfigurasi terbaik, dilakukan eksperimen terhadap 24 skenario kombinasi hyperparameter yang terdiri dari epoch (5 dan 10), batch size (16 dan 32), serta dropout (0,2; 0,4; dan 0,6). Evaluasi setiap skenario dilakukan menggunakan metode *5-Fold Cross Validation*. Sebelum proses pencarian hyperparameter, dilakukan penanganan ketidakseimbangan kelas pada model *baseline* menggunakan kombinasi *oversampling* dan *focal loss*. Hasil menunjukkan bahwa pendekatan tersebut mampu meningkatkan performa klasifikasi secara konsisten pada seluruh aspek, sehingga digunakan sebagai dasar dalam eksperimen pemilihan hyperparameter.

3.7. Evaluasi Model

Evaluasi menggunakan *5-Fold Cross Validation* bertujuan untuk menentukan konfigurasi hyperparameter terbaik pada masing-masing aspek. Pemilihan model dilakukan berdasarkan nilai *mean F1-score* dan standar deviasi F1-score. Berikut konfigurasi *hyperparameter* model terbaik pada masing-masing aspek.

Tabel 4. Hasil Evaluasi 5-Fold Cross Validation

Model	Konfigurasi	Mean F1-Score	STD F1
Performa Terbaik Model IndoBERT			
Tampilan	Epoch : 10 Batch Size : 32 Dropout : 0.2	0.86	0.023
Fitur dan Performa	Epoch : 10 Batch Size : 16 Dropout : 0.2	0.83	0.006
Layanan	Epoch : 5 Batch Size : 32 Dropout : 0.2	0.74	0.026
Performa Terbaik Model mBERT			
Tampilan	Epoch : 10 Batch Size : 16 Dropout : 0.2	0.80	0.030
Fitur dan Performa	Epoch : 10 Batch Size : 32 Dropout : 0.2	0.79	0.013
Layanan	Epoch : 5 Batch Size : 16 Dropout : 0.2	0.67	0.031

Berdasarkan hasil evaluasi *5-Fold Cross Validation* pada Tabel 4, diperoleh nilai *F1-score* tertinggi dengan standar deviasi yang relatif rendah. Hal ini menunjukkan bahwa performa model konsisten dan stabil pada setiap *fold* pengujian. Hasil ini juga menunjukkan bahwa IndoBERT secara konsisten mengungguli mBERT pada seluruh aspek yang dianalisis.

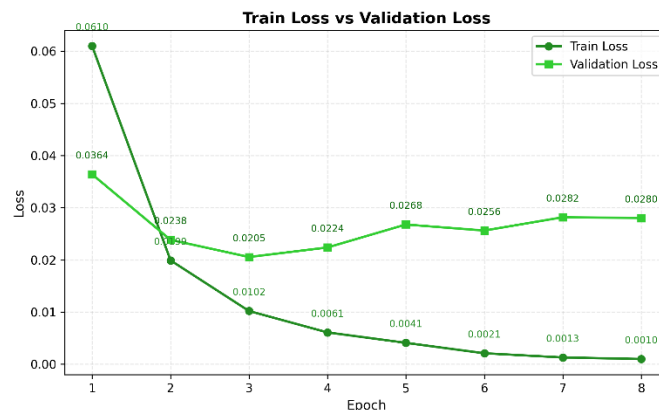
Setelah diperoleh konfigurasi terbaik dari tahap *5-Fold Cross Validation*, dilakukan final training menggunakan seluruh data pelatihan. Pada tahap ini, data pelatihan kembali dibagi menjadi data latih dan data validasi dengan rasio 90:10. Data validasi digunakan untuk memantau performa model selama pelatihan serta menerapkan mekanisme *early stopping* dengan nilai *patience* sebesar 2 untuk mencegah *overfitting*. Berikut adalah tabel hasil evaluasi *final training*.

Tabel 5. Hasil Evaluasi Final Training

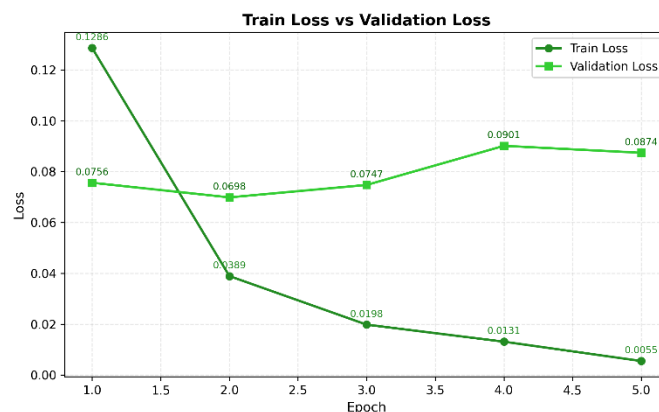
Aspek	Accuracy	Precision	Recall	F1-Score
Tampilan	0.97	0.84	0.83	0.83
Fitur dan Performa	0.95	0.91	0.87	0.89
Layanan	0.89	0.83	0.87	0.85

Tabel 5 menunjukkan hasil evaluasi final training pada masing-masing aspek. Aspek Fitur dan Performa memperoleh performa terbaik dengan F1-score sebesar

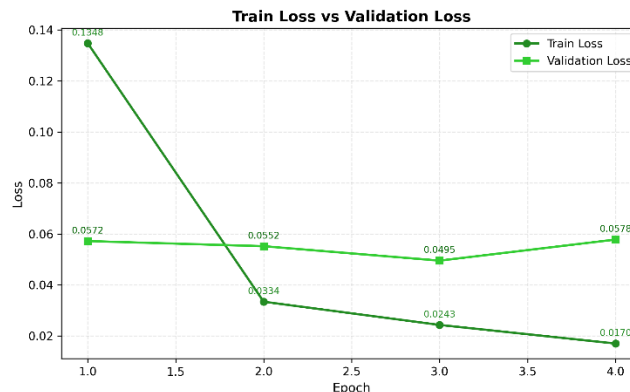
0,89, diikuti aspek Layanan sebesar 0,85 dan aspek Tampilan sebesar 0,83. Untuk mengevaluasi proses pembelajaran model selama tahap *final training*, dilakukan pemantauan terhadap nilai train loss dan validation loss pada setiap epoch. Grafik ini digunakan untuk mengamati kemampuan model dalam mempelajari data pelatihan sekaligus mengevaluasi performanya pada data validasi. Penurunan *train loss* mencerminkan peningkatan kemampuan model dalam mempelajari pola data, sedangkan *validation loss* digunakan untuk mengidentifikasi potensi *overfitting*.



Gambar 3. Grafik Train Loss vs Validation Loss (Tampilan)



Gambar 4. Grafik Train Loss vs Validation Loss (Fitur dan Performa)



Gambar 5. Grafik Train Loss vs Validation Loss (Layanan)

Gambar 6, 7, dan 8 menunjukkan perkembangan nilai *train loss* dan *validation loss* selama proses *final training* pada masing-masing aspek. Secara umum, nilai *train loss* mengalami penurunan secara konsisten pada seluruh aspek, yang mengindikasikan bahwa model semakin mampu mempelajari pola dari data pelatihan. Sementara itu, *validation loss* mencapai nilai terendah pada epoch ke-2 hingga ke-3 sebelum mengalami sedikit peningkatan atau fluktuasi pada epoch berikutnya. Kondisi ini mengindikasikan bahwa model telah mencapai performa optimal dengan kemampuan generalisasi yang baik terhadap data validasi. Oleh karena itu, model pada *epoch* dengan nilai *validation loss* terendah dipilih sebagai model terbaik melalui mekanisme *early stopping* untuk digunakan pada tahap pengujian.

Evaluasi akhir dilakukan pada data uji yang tidak digunakan selama proses pelatihan sehingga hasil yang diperoleh mencerminkan kemampuan generalisasi model. Berikut adalah hasil evaluasi pada data test.

Tabel 6. Hasil Evaluasi Data Test

Aspek	Accuracy	Precision	Recall	F1-Score
Tampilan	0.96	0.86	0.78	0.80
Fitur dan Performa	0.92	0.86	0.81	0.83
Layanan	0.91	0.80	0.84	0.81

Tabel 6 menunjukkan bahwa model memperoleh performa yang baik pada seluruh aspek. Aspek fitur dan performa memperoleh performa terbaik dengan F1-score sebesar 0,83, diikuti aspek layanan sebesar 0,81 dan aspek tampilan sebesar 0,80. Hasil ini menunjukkan bahwa model mampu melakukan klasifikasi sentimen berbasis aspek secara konsisten pada data yang belum pernah digunakan selama proses pelatihan.

Tabel 7. Classification Report

Aspek	Sentimen	Precision	Recall	F1-Score
Tampilan	Positif	0.70	0.82	0.76
	Negatif	0.89	0.53	0.67
	None	0.98	0.98	0.98
Fitur dan Performa	Positif	0.89	0.92	0.90
	Negatif	0.94	0.98	0.96
	Netral	0.70	0.63	0.67
	None	0.90	0.71	0.80
Layanan	Positif	0.97	0.92	0.94
	Negatif	0.80	0.73	0.77
	Netral	0.50	0.75	0.60
	None	0.93	0.95	0.94

Berdasarkan *classification report* pada tabel 7, model menunjukkan performa yang baik pada seluruh aspek, terutama pada kelas none dengan F1-score mencapai 0,98. Performa terbaik diperoleh pada sentimen negatif aspek Fitur dan Performa (F1-score 0,96) serta sentimen positif aspek Layanan (F1-score 0,94). Nilai F1-score yang relatif lebih rendah ditemukan pada kelas netral karena jumlah datanya lebih sedikit.

KESIMPULAN

Penelitian ini berhasil menerapkan analisis sentimen berbasis aspek pada ulasan JConnect Mobile menggunakan metode *fine-tuning* IndoBERT. Berdasarkan hasil evaluasi yang telah dilakukan, IndoBERT menunjukkan kinerja yang lebih unggul dibandingkan mBERT pada seluruh aspek sentimen. Dengan demikian, IndoBERT ditetapkan sebagai model terbaik. Hasil tersebut menunjukkan bahwa *fine-tuning* IndoBERT efektif untuk mengklasifikasikan sentimen berbasis aspek pada ulasan pengguna aplikasi berbahasa Indonesia. Kontribusi penelitian ini adalah penerapan analisis sentimen berbasis aspek sehingga sentimen dapat dianalisis secara lebih spesifik berdasarkan aspek yang dibahas dalam ulasan JConnect Mobile menggunakan IndoBERT dengan mengombinasikan data dari Google Play Store dan App Store, serta mengevaluasi performa model melalui *hyperparameter tuning* dan *5-Fold Cross Validation*. Secara praktis, hasil penelitian ini dapat membantu Bank Jatim dalam mengidentifikasi aspek yang paling banyak memperoleh tanggapan positif maupun negatif dari pengguna. Informasi tersebut dapat dimanfaatkan sebagai bahan evaluasi untuk meningkatkan kualitas tampilan, fitur dan performa, serta layanan JConnect Mobile secara lebih terarah. Penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam, mengembangkan aspek sentimen yang lebih spesifik, serta membandingkan performa IndoBERT dengan model Transformer yang lebih mutakhir, seperti XLM-RoBERTa maupun *Large Language Model* (LLM).

DAFTAR PUSTAKA

- [1] A. Kurniawan, A. Rahayu, and L. A. Wibowo, "Pengaruh Transformasi Digital Terhadap Kinerja Bank Pembangunan Daerah di Indonesia," *Jurnal Ilmu Keuangan Dan Perbankan (JIKA)*, vol. 10, no. 2, pp. 158–181, 2021.
- [2] Bank Jatim, "'JConnect', Solusi Kemudahan Akses Layanan Digital Perbankan dari BJTM," PT Bank Pembangunan Daerah Jawa Timur Tbk.
- [3] N. Khasanah, F. M. Amin, and A. Permadi, "The Effect of User Experience and Usability on User Satisfaction and Continuance Intention in the JConnect Mobile Application," *TIERS Information Technology Journal*, vol. 4, no. 2, pp. 171–178, 2023.
- [4] Google Play Store, "Jconnect Mobile." Accessed: Dec. 11, 2025. [Online]. Available: <https://play.google.com/store/apps/details?id=com.dwidasa.bjtm.mb.android&hl=en>
- [5] Bank Pembangunan Daerah Jawa Timur Tbk., "Laporan Tahunan 2024 Mencapai Kinerja Unggul melalui Sinergi dan Kolaborasi," Bank Pembangunan Daerah Jawa Timur Tbk. Accessed: Dec. 07, 2025. [Online]. Available: https://www.bankjatim.co.id/files/iru/bahan_rups/laporan_tahunan/2024/laporan_tahunan_bjtm_2024.pdf
- [6] N. Hidayati, F. Hamami, and R. Y. Fa'rifah, "Aspect-Based Sentiment Analysis On FLIP Application Reviews (Play Store) Using Support Vector Machine (SVM) Algorithm," *Journal of Informatics and Telecommunication Engineering*, vol. 7, no. 1, pp. 183–197, 2023.
- [7] C. Apriyadi and S. Styawati, "Sentiment Analysis of Cyber Attacks in Bank Syariah Indonesia Using SVM and Indobert Method," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 819–838, 2025.
- [8] A. Aprinando, P. Simarmata, and T. B. Sasongko, "Sentiment Analysis on BRImo Application Reviews Using IndoBERT," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [9] A. R. Putra and D. E. Ratnawati, "Analisis Sentimen Berbasis Aspek pada Aplikasi Mobile Menggunakan Naïve Bayes berdasarkan Ulasan Pengguna Playstore (Studi Kasus: Jconnect Mobile)," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 293–300, 2025.
- [10] K. Darmanto and S. Sugiono, "ANALISIS SENTIMEN ULASAN APLIKASI JCONNECT MOBILE DI GOOGLE PLAY STORY MENGGUNAKAN ALGORITMA MULTINOMIAL NAÏVE BAYES DAN SUPPORT VECTORE MACHINE," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 5, pp. 8357–8364, 2025.
- [11] P. Zhang, X. Huang, Y. Wang, C. Jiang, S. He, and H. Wang, "Semantic similarity computing model based on multi model fine-grained nonlinear fusion," *IEEE Access*, vol. 9, pp. 8433–8443, 2021.
- [12] A. Jazuli, Widowati, and R. Kusumaningrum, "Optimizing Aspect-Based Sentiment Analysis Using BERT for Comprehensive Analysis of Indonesian Student Feedback," *Applied Sciences*, vol. 15, no. 1, p. 172, 2024.

- [13]H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment analysis in Indonesian healthcare applications using IndoBERT approach," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023.
- [14]R. I. Perwira, V. A. Permadi, D. I. Purnamasari, and R. P. Agusdin, "Domain-Specific Fine-Tuning of IndoBERT for Aspect-Based Sentiment Analysis in Indonesian Travel User-Generated Content," *Journal of Information Systems Engineering and Business Intelligence*, vol. 11, no. 1, pp. 30–40, 2025.
- [15]G. Marzi, M. Balzano, and D. Marchiori, "K-Alpha calculator–krippendorff's alpha calculator: a user-friendly tool for computing krippendorff's alpha inter-rater reliability coefficient," *MethodsX*, vol. 12, p. 102545, 2024.