

Optimasi Klasifikasi Keluhan Pengguna *E-Wallet* Berbasis *Deep Learning* IndoBERT untuk Mendukung Manajemen Layanan TI

Sarah Astiti^{1*}, Yeyi Gusla Nengsih², Khairunnisa Samosir³, Darmansah⁴

¹Program Studi Sistem Informasi, Universitas Telkom, Kampus Purwokerto, Indonesia

²Program Studi Sistem Informasi, Sekolah Tinggi Teknologi Industri Padang, Indonesia

³Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara, Indonesia

⁴Program Studi Sistem Informasi, Fakultas Teknik dan Komputer, Universitas Putera Batam, Indonesia

¹sarahas@telkomuniversity.ac.id; ²yeyigusla22@gmail.com; ³khairunnisasamosir01@gmail.com; ⁴darmansah071@gmail.com

Abstrak.

Tujuan: Peningkatan penggunaan *e-wallet* di Indonesia menyebabkan jumlah keluhan pengguna terkait layanan digital semakin beragam, sehingga diperlukan metode klasifikasi yang mampu mengelompokkan keluhan secara cepat dan akurat untuk mendukung Manajemen Layanan TI. Penelitian ini bertujuan mengoptimalkan klasifikasi keluhan pengguna *e-wallet* menggunakan IndoBERT serta mengkaji pemanfaatannya dalam mendukung proses *service desk* dan *incident management*.

Metode/Design/Pendekatan: Penelitian menggunakan pendekatan kuantitatif dan data penelitian diperoleh dari ulasan pengguna aplikasi *e-wallet* pada platform *Google Play Store* yang terdiri atas tujuh kategori, yaitu Transaksi, Top Up, Saldo, Login, Verifikasi, Aplikasi, dan Layanan Pelanggan, dengan masing-masing kategori sebanyak 100 data. Dataset dibagi secara *stratified* menjadi 560 data training (80%), 70 data validation (10%), dan 70 data testing (10%). Tahapan penelitian meliputi *preprocessing*, pelabelan, tokenisasi IndoBERT, *fine-tuning*, optimasi *hyperparameter*, serta evaluasi menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

Hasil/Temuan: Hasil eksperimen menunjukkan IndoBERT Baseline memperoleh *accuracy* 98,57%, sedangkan IndoBERT Optimasi mencapai 100% pada *accuracy*, *precision*, *recall*, dan *F1-score*. *Confusion Matrix* menunjukkan seluruh 70 data testing berhasil diklasifikasikan dengan benar.

Kebaharuan/Originalitas/Nilai: Kontribusi penelitian ini terletak pada pengoptimalan model IndoBERT untuk mengklasifikasikan keluhan *e-wallet* berdasarkan kategori permasalahan layanan TI dan pemanfaatan hasil klasifikasi tersebut untuk mendukung proses Manajemen Layanan TI, khususnya kategorisasi keluhan, pengelompokan insiden, dan *service desk*.

Keywords: *E-Wallet, IndoBERT, Complaint Classification, Deep Learning, IT Service Management.*

Abstract.

Purpose: The increasing use of *e-wallets* in Indonesia has led to an increasingly diverse number of user complaints related to digital services, necessitating a classification method capable of quickly and accurately grouping complaints to support IT Service Management. This study aims to optimize the classification of *e-wallet* user complaints using IndoBERT and examine its use in supporting *service desk* and *incident management* processes.

Method/Design/Approach: The study used a quantitative approach and research data were obtained from user reviews of *e-wallet* applications on the *Google Play Store* platform, consisting of seven categories: Transactions, Top Up, Balance, Login, Verification, Applications, and Customer Service, with 100 data points each. The dataset was stratified into 560 training data points (80%), 70 validation data points (10%), and 70 testing data points (10%). The research stages included *preprocessing*, labeling, IndoBERT tokenization, *fine-tuning*, *hyperparameter* optimization, and evaluation using *accuracy*, *precision*, *recall*, *F1-score*, and *confusion matrix*.

Results/Findings: Experimental results show that IndoBERT Baseline achieved 98.57% accuracy, while IndoBERT Optimization achieved 100% accuracy, precision, recall, and F1-score. The Confusion Matrix shows that all 70 test data sets were correctly classified.

Novelty/Originality/Value: The contribution of this research lies in optimizing the IndoBERT model to classify *e-wallet* complaints based on IT service problem categories and utilizing the classification results to support IT Service Management processes, specifically complaint categorization, incident grouping, and *service desk*.

Keywords: *E-Wallet, IndoBERT, Complaint Classification, Deep Learning, IT Service Management.*

Article history:

Received, 2026-09-07

Revised, 2026-09-28

Accepted, 2026-10-14

^{1*}Corresponding author.

Sarah Astiti.

Email addresses: sarahas@telkomuniversity.ac.id

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

PENDAHULUAN

Pengembangan teknologi digital telah mengubah cara orang Indonesia bertransaksi [1], [2], [3], [4]. Salah satu contohnya adalah e-wallet, yang memudahkan transfer dana, pembayaran, membeli barang digital, dan berbagai transaksi lainnya [5], [6]. Selain itu, data Bank Indonesia menunjukkan peningkatan penggunaan layanan pembayaran digital. Mereka memperkirakan volume transaksi pembayaran digital akan meningkat 36,1% pada tahun 2024, dengan pertumbuhan transaksi QRIS sebesar 175,2% per tahun [7], [8]. Dengan peningkatan ini, layanan pembayaran digital telah menjadi bagian penting dari kehidupan masyarakat, dan penyedia e-wallet diharuskan untuk menyediakan layanan TI yang cepat, stabil, dan dapat diandalkan. Peningkatan jumlah pengguna dan volume transaksi juga berpotensi meningkatkan jumlah permasalahan yang dialami pengguna. Beberapa contoh masalah yang sering muncul yaitu kegagalan top-up, masalah login dan verifikasi, kesalahan aplikasi, dan masalah layanan [9], [10]. Data keluhan tersebut sebenarnya memiliki informasi yang penting bagi penyedia layanan karena dapat menggambarkan permasalahan yang terjadi secara langsung dari perspektif pengguna. Namun, apabila jumlah keluhan semakin besar dan proses pengelompokan masih dilakukan secara manual, diperlukan waktu dan sumber daya yang lebih banyak. Selain itu, perbedaan cara pengguna menyampaikan keluhan dapat menyebabkan kesulitan dalam menentukan kategori permasalahan secara konsisten.

Proses klasifikasi juga menghadapi karakteristik bahasa ulasan pengguna. Untuk menyelesaikan masalah yang sama, pengguna sering menggunakan bahasa yang tidak formal, singkatan, kesalahan pengetikan, slang, dan berbagai istilah. Sebagai contoh, keluhan "saldo kepotong tapi transaksi gagal" dan "uang sudah terpotong pembayaran tidak masuk" dapat menunjukkan masalah yang sama. Karena keadaan ini, metode klasifikasi berbasis frekuensi kata seperti TF-IDF tidak dapat memahami konteks setiap kalimat dengan baik. Akibatnya, dibutuhkan teknik yang mampu memahami lebih baik hubungan antar kata dan konteks bahasa. Untuk menganalisis ulasan aplikasi digital, penelitian sebelumnya telah menggunakan *machine learning* dan *deep learning*. TF-IDF dikombinasikan dengan algoritma seperti *Support Vector Machine*, *Naive Bayes*, dan *Random Forest*. Dalam penelitian tambahan, LSTM dan BiLSTM digunakan untuk meningkatkan kemampuan klasifikasi teks berbahasa Indonesia. Meskipun teknik ini dapat bekerja dengan baik, model masih terbatas dalam memahami konteks kata yang luas. Pengembangan model Transformer, terutama *Bidirectional Encoder Representations from Transformers* (BERT), menghasilkan pemahaman teks yang lebih baik [9], [11], [12], [13], [14].

IndoBERT adalah model Bahasa Indonesia yang memiliki kemampuan untuk menampilkan teks berdasarkan konteks, yang memungkinkannya menangani berbagai bahasa yang ditemukan dalam ulasan pengguna [15], [16], [17]. Penelitian sebelumnya menggunakan 5.000 data ulasan untuk menganalisis aplikasi *e-wallet* dan menggunakan IndoBERT untuk menganalisis 4.669 ulasan setelah proses *preprocessing* [18], [19], [20]. Hasilnya menunjukkan bahwa IndoBERT memiliki akurasi sebesar 84,36% dalam klasifikasi sentimen, yang menunjukkan kemampuan yang luar biasa untuk mengolah teks ulasan *e-wallet*. Selain itu, penelitian tersebut menemukan masalah seperti layanan pelanggan, kesalahan aplikasi, login dan verifikasi, saldo, dan transaksi yang gagal. Meskipun demikian, penelitian tersebut masih berfokus pada klasifikasi sentimen, sehingga hasilnya hanya menunjukkan apakah suatu ulasan memiliki sentimen positif, negatif, atau netral. Informasi tersebut belum secara langsung menunjukkan jenis permasalahan layanan TI yang harus ditangani. Sebagai contoh, dua ulasan yang sama-sama memiliki sentimen negatif dapat berasal dari permasalahan transaksi gagal dan masalah login. Kedua permasalahan tersebut membutuhkan proses penanganan yang berbeda. Dengan demikian, terdapat kesenjangan penelitian berupa kebutuhan untuk mengembangkan klasifikasi yang tidak hanya mengidentifikasi sentimen, tetapi juga mengelompokkan keluhan berdasarkan jenis permasalahan layanan.

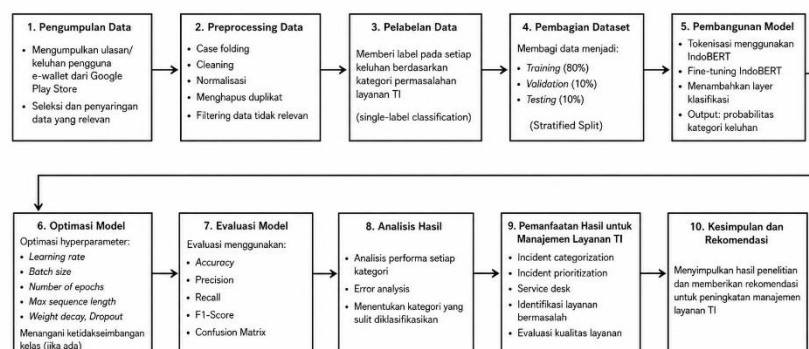
Penelitian sebelumnya telah menunjukkan bahwa IndoBERT mampu digunakan untuk mengklasifikasikan sentimen pada ulasan pengguna *e-wallet*, tetapi belum secara khusus mengoptimalkan IndoBERT untuk mengklasifikasikan keluhan berdasarkan jenis permasalahan layanan TI yang dapat mendukung kebutuhan manajemen layanan. Oleh karena itu, penelitian ini mengisi kesenjangan tersebut dengan mengembangkan dan mengoptimalkan model IndoBERT untuk mengelompokkan ulasan pengguna *e-wallet* ke dalam kategori permasalahan layanan TI. Proses penelitian mencakup pelabelan data berdasarkan kategori keluhan, *fine-tuning* IndoBERT, dan optimasi parameter model, kemudian mengevaluasi performanya menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Dengan pendekatan tersebut, penelitian ini tidak hanya mengidentifikasi sentimen pengguna, tetapi menghasilkan informasi mengenai jenis permasalahan yang dapat digunakan sebagai dasar untuk mendukung proses pengelolaan layanan TI.

Dalam konteks *IT Service Management* (ITSM), klasifikasi keluhan berdasarkan kategori permasalahan memiliki nilai yang lebih praktis. Hasil klasifikasi dapat membantu proses *incident categorization*, *incident management*, dan *service desk* dengan memberikan informasi mengenai jenis gangguan yang paling sering dialami pengguna. Informasi tersebut juga dapat digunakan untuk mengidentifikasi layanan yang membutuhkan perhatian lebih besar serta menjadi bahan evaluasi dalam meningkatkan kualitas layanan TI [21], [22], [23]. Dengan demikian, data

keluhan pengguna yang sebelumnya tidak terstruktur dapat diubah menjadi informasi yang lebih terorganisasi dan mendukung pengambilan keputusan. Berdasarkan permasalahan tersebut, penelitian ini mengusulkan optimasi klasifikasi keluhan pengguna *e-wallet* berbasis deep learning IndoBERT untuk mendukung manajemen layanan TI. Penelitian dilakukan melalui tahapan pengumpulan data ulasan, *preprocessing*, pelabelan kategori keluhan, *fine-tuning* IndoBERT, serta optimasi parameter model. Performa model akan dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Penggunaan beberapa metrik tersebut diperlukan agar kemampuan model tidak hanya dinilai berdasarkan jumlah prediksi yang benar, tetapi juga berdasarkan kemampuannya mengenali setiap kategori keluhan. Penelitian ini berfokus pada penggunaan dan pengoptimalan IndoBERT untuk mengklasifikasikan jenis keluhan *e-wallet* yang dimaksudkan untuk memenuhi kebutuhan manajemen layanan TI, bukan hanya untuk analisis sentimen. Hasil klasifikasi ini diharapkan dapat memberikan pemetaan jenis masalah pengguna yang lebih akurat, yang akan membantu proses identifikasi insiden, pengelompokan keluhan, penentuan prioritas penanganan, dan evaluasi layanan. Oleh karena itu, penelitian ini tidak hanya membantu penggunaan NLP Bahasa Indonesia, tetapi juga membantu menggunakan deep learning untuk meningkatkan efisiensi manajemen layanan TI pada layanan *e-wallet*.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan menerapkan metode *Natural Language Processing* (NLP) dan *deep learning*. Penelitian bertujuan membangun dan mengoptimalkan model IndoBERT untuk mengklasifikasikan keluhan pengguna *e-wallet* berdasarkan kategori permasalahan layanan TI. Tahapan penelitian dilakukan secara sistematis mulai dari pengumpulan data keluhan pengguna, *preprocessing* data teks, pelabelan data, pembagian dataset, pembangunan model IndoBERT, optimasi model, evaluasi performa, hingga analisis hasil klasifikasi untuk mendukung manajemen layanan TI. Berikut ini Adalah tahapan dari penelitian ini:



Gambar 1 Tahapan Penelitian

Penelitian dimulai dengan pengumpulan data ulasan pengguna *e-wallet*, *preprocessing*, pelabelan, pembagian dataset, pembuatan dan penyempurnaan model IndoBERT, evaluasi, analisis hasil, dan pemanfaatan hasil klasifikasi untuk mendukung manajemen layanan TI. Data penelitian dikumpulkan dari ulasan pengguna aplikasi *e-wallet* di *Google Play Store*. Ulasan tersebut berupa teks yang berkaitan dengan pengalaman dan masalah pengguna saat menggunakan layanan *e-wallet*. Ulasan yang tidak relevan, seperti iklan, spam, komentar kosong, atau komentar yang tidak memberikan informasi tentang masalah layanan, kemudian dipilih dan digunakan sebagai dataset penelitian. Tahap berikutnya adalah *preprocessing* untuk meningkatkan kualitas data teks sebelum digunakan dalam proses pemodelan. Proses ini meliputi *case folding* untuk menyeragamkan huruf, *cleaning* untuk menghilangkan karakter yang tidak diperlukan, *normalisasi* kata tidak baku, penghapusan data duplikat, serta penyaringan data yang tidak relevan.

Setelah proses *preprocessing* selesai, setiap ulasan dilabelkan berdasarkan kategori masalah layanan TI. Kategori ini dapat mencakup masalah aplikasi, login, verifikasi, transaksi, saldo, top-up, dan layanan pelanggan. Pelabelan dilakukan berdasarkan isi dan konteks keluhan. Pedoman pelabelan dibuat untuk memastikan bahwa setiap data memiliki kategori yang seragam. Dalam kasus di mana penelitian menggunakan klasifikasi satu label, setiap ulasan hanya diberikan satu kategori utama yang paling tepat mencerminkan masalah yang dihadapi pengguna. Dataset yang telah diberi label kemudian dibagi menjadi data pembelajaran, validasi, dan pengujian dengan proporsi 80%, 10%, dan 10%. Pembagian dilakukan dengan metode pembagian *stratified* untuk menjaga proporsi

masing-masing kategori pada ketiga dataset. Pada tahap pemodelan, teks keluhan diubah menjadi representasi token menggunakan *tokenizer* IndoBERT. Token tersebut kemudian diproses oleh arsitektur Transformer untuk menghasilkan representasi teks yang mempertimbangkan konteks setiap kata dalam kalimat. Representasi tersebut diteruskan ke lapisan klasifikasi untuk menentukan kategori keluhan. Probabilitas setiap kategori diperoleh melalui fungsi Softmax yang dirumuskan sebagai persamaan (1) berikut:

$$p(y_i|x) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (1)$$

Keterangan:

$p(y_i|x)$ = merupakan probabilitas data x termasuk ke dalam kelas i

z_i = merupakan nilai keluaran (logit) untuk kelas i

K = merupakan jumlah kategori keluhan

Kategori dengan probabilitas tertinggi digunakan sebagai hasil prediksi model. Kemudian Proses pembelajaran model dilakukan melalui fine-tuning IndoBERT menggunakan data training. Selama proses tersebut, parameter model disesuaikan dengan karakteristik data keluhan *e-wallet*. Fungsi loss yang digunakan dapat berupa *Cross-Entropy Loss* untuk mengukur perbedaan antara label sebenarnya dan prediksi model. Persamaan *Cross-Entropy Loss* dapat dituliskan sebagai persamaan (2) berikut:

$$L = - \sum_{i=1}^K y_i \log(y_i) \quad (2)$$

Keterangan:

L = merupakan nilai loss

y_i = merupakan label sebenarnya

$\log y_i$ = merupakan probabilitas hasil prediksi model pada kelas ke - i

Semakin kecil nilai loss, semakin baik kemampuan model dalam menyesuaikan prediksi dengan label sebenarnya. Tahap selanjutnya adalah optimasi model untuk memperoleh konfigurasi IndoBERT dengan performa terbaik. Optimasi dilakukan dengan menguji beberapa *hyperparameter*, seperti *learning rate*, *batch size*, *number of epochs*, *maximum sequence length*, *weight decay*, dan *dropout*. Pemilihan model terbaik dilakukan berdasarkan hasil pada validation set, terutama dengan mempertimbangkan nilai *macro F1-score* agar setiap kategori memiliki pengaruh yang seimbang terhadap penilaian model. Setelah proses optimasi selesai, model terbaik diuji menggunakan data testing. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Selain evaluasi menggunakan metrik numerik, digunakan *confusion matrix* untuk melihat pola kesalahan klasifikasi pada setiap kategori. Analisis tersebut dapat menunjukkan kategori keluhan yang berhasil dikenali dengan baik serta kategori yang sering tertukar. Selanjutnya dilakukan error analysis terhadap beberapa contoh prediksi yang salah untuk mengetahui penyebab kesalahan, misalnya karena penggunaan istilah yang serupa antara kategori transaksi dan saldo, penggunaan bahasa informal, singkatan, atau adanya lebih dari satu permasalahan dalam satu ulasan. Kemudian, hasil klasifikasi dianalisis untuk mengetahui distribusi dan dominasi masing-masing kategori keluhan pengguna. Informasi ini kemudian dikaitkan dengan kebutuhan manajemen layanan TI. Kategori keluhan dengan jumlah tinggi dapat menunjukkan bahwa layanan yang memerlukan perhatian lebih besar. Selain itu, hasil klasifikasi dapat digunakan untuk mendukung proses kategorisasi keluhan, prioritasasi keluhan, meja layanan, identifikasi layanan bermasalah, dan tujuan lainnya.

HASIL DAN PEMBAHASAN

Data penelitian diperoleh dari ulasan pengguna aplikasi *e-wallet* pada platform *Google Play Store*. Data tersebut dikelompokkan ke dalam tujuh kategori masalah layanan TI: transaksi, top up, saldo, login, verifikasi, aplikasi, dan layanan pelanggan. Setiap kategori memiliki jumlah data yang sama, atau 14,29% dari seluruh dataset. Karena tidak ada kategori dengan jumlah data yang jauh lebih besar daripada kategori lainnya, distribusi data yang seimbang ini membantu dalam proses pelatihan model. Kondisi ini juga dapat mengurangi kecenderungan model untuk memprediksi lebih banyak tentang kelas tertentu. Berikut Adalah dataset pada penelitian ini yang bisa dilihat pada tabel 1 dibawah ini:

Tabel 1. Data Set

Kategori Keluhan	Jumlah	Persentase
Transaksi	100	14,29%
Top Up	100	14,29%
Saldo	100	14,29%
Login	100	14,29%
Verifikasi	100	14,29%
Aplikasi	100	14,29%
Layanan Pelanggan	100	14,29%
Total	700	100%

Tabel tersebut menunjukkan bahwa dataset memiliki distribusi kelas yang seimbang. Ini berbeda dengan data keluhan nyata, di mana kategori tertentu biasanya memiliki lebih banyak keluhan daripada kategori lainnya. Oleh karena itu, pada tahap penelitian selanjutnya, karakteristik distribusi data nyata harus diperhatikan agar model yang dihasilkan dapat mewakili kondisi operasional sebenarnya. Dataset kemudian dibagi menjadi data training, validation, dan testing dengan rasio 80:10:10. Karena setiap kategori terdiri dari 100 data, maka masing-masing kategori memperoleh 80 data untuk training, 10 data untuk validation, dan 10 data untuk testing. Dengan demikian, keseluruhan dataset terdiri dari 560 data training, 70 data validation, dan 70 data testing. Berikut ini merupakan pembagian data set yang bisa dilihat pada tabel 2 dibawah ini:

Tabel 2. Hasil *Preprocessing* Data

Kategori	Training	Validation	Testing	Total
Transaksi	80	10	10	100
Top Up	80	10	10	100
Saldo	80	10	10	100
Login	80	10	10	100
Verifikasi	80	10	10	100
Aplikasi	80	10	10	100
Layanan Pelanggan	80	10	10	100
Total	560	70	70	700

Hasil eksperimen menunjukkan bahwa TF-IDF + *Logistic Regression* dan IndoBERT Optimasi sama-sama memperoleh *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100%. Temuan ini menunjukkan bahwa, metode TF-IDF + *Logistic Regression* sudah mampu membedakan kategori keluhan dengan sangat baik. Sementara itu, IndoBERT baseline memperoleh *accuracy* 98,57%, *precision* 98,61%, *recall* 98,57%, dan *F1-score* 98,57%, kemudian meningkat menjadi 100% setelah dilakukan optimasi. Dengan demikian, hasil tersebut menunjukkan bahwa optimasi konfigurasi pelatihan dapat meningkatkan performa IndoBERT pada dataset yang digunakan, tetapi tidak menunjukkan bahwa IndoBERT secara mutlak lebih unggul dibandingkan TF-IDF + *Logistic Regression*. Optimasi IndoBERT dilakukan dengan mempertimbangkan beberapa hyperparameter, yaitu *learning rate*, *batch size*, *epoch*, *maximum sequence length*, *weight decay*, dan *dropout*. Beberapa kombinasi konfigurasi dibandingkan berdasarkan performa pada data validasi, kemudian konfigurasi dengan hasil terbaik dipilih sebagai konfigurasi optimal. Konfigurasi terpilih menggunakan *learning rate* 2e-5, *batch size* 16, *epoch* 5, *maximum sequence length* 128, dan *weight decay* 0,01, sedangkan nilai *dropout* mengikuti konfigurasi yang menghasilkan performa validasi terbaik. Hasil ini menunjukkan bahwa perubahan konfigurasi pelatihan berkontribusi terhadap peningkatan performa IndoBERT dari 98,57% menjadi 100% pada dataset penelitian. Meskipun memperoleh nilai 100%, hasil tersebut perlu diinterpretasikan sebagai performa pada dataset yang terbatas dan tidak dapat langsung dianggap sebagai bukti bahwa model akan memiliki performa yang sama pada kondisi operasional. Nilai tersebut dapat dipengaruhi oleh karakteristik dataset, seperti jumlah data, distribusi kelas, tingkat kemiripan data, dan kejelasan perbedaan antar kategori. Kesamaan hasil antara TF-IDF + *Logistic Regression* dan IndoBERT Optimasi juga menunjukkan bahwa model yang lebih kompleks belum tentu menghasilkan peningkatan metrik pada dataset tertentu. Oleh karena itu, keunggulan representasi kontekstual IndoBERT perlu diuji lebih lanjut menggunakan data yang lebih besar dan beragam untuk mengetahui ketahanannya terhadap variasi bahasa dan pola keluhan pengguna yang lebih kompleks. Berikut ini adalah hasil dari Perbandingan *Baseline* dan IndoBERT Optimasi yang bisa dilihat pada tabel 3 berikut:

Tabel 3. Hasil Perbandingan *Baseline* dan IndoBERT Optimasi

Eksperimen	Learning Rate	Batch Size	Epoch	Max Length	Weight Decay	Accuracy	Precision	Recall	F1-Score
Baseline TF-IDF + <i>Logistic Regression</i>	-	-	-	-	-	100.00%	100.00%	100.00%	100.00%
IndoBERT Baseline	2e-5	16	3	128	0.00	98.57%	98.61%	98.57%	98.57%
IndoBERT Optimasi	2e-5	16	5	128	0.01	100.00%	100.00%	100.00%	100.00%

Kemudian berdasarkan hasil *Classification Report*, seluruh kategori memperoleh nilai *precision*, *recall*, dan *F1-score* sebesar 1,00 (100%) dengan masing-masing support sebanyak 10 data. Hal ini menunjukkan bahwa model IndoBERT optimasi mampu mengklasifikasikan seluruh data testing pada kategori Transaksi, Top Up, Saldo, Login, Verifikasi, Aplikasi, dan Layanan Pelanggan dengan tepat tanpa kesalahan. Nilai *macro average* dan *weighted average* sebesar 1,00 juga menunjukkan bahwa performa model konsisten pada seluruh kategori. Hasil tersebut mengindikasikan kemampuan klasifikasi yang sangat baik pada dataset penelitian, meskipun perlu diperhatikan bahwa dataset yang digunakan terbatas sehingga hasil pada data keluhan pengguna nyata dapat berbeda. Berikut hasil *Classification Report* yang bisa dilihat pada tabel 4 berikut ini:

Tabel 4. Hasil *Classification Report*

Kategori	Precision	Recall	F1-Score	Support
Transaksi	1.00	1.00	1.00	10
Top Up	1.00	1.00	1.00	10
Saldo	1.00	1.00	1.00	10
Login	1.00	1.00	1.00	10
Verifikasi	1.00	1.00	1.00	10
Aplikasi	1.00	1.00	1.00	10
Layanan Pelanggan	1.00	1.00	1.00	10
Macro Average	1.00	1.00	1.00	70
Weighted Average	1.00	1.00	1.00	70

Berdasarkan *Confusion Matrix* IndoBERT Baseline, dari 70 data testing terdapat 69 data yang berhasil diklasifikasikan dengan benar dan 1 data mengalami kesalahan klasifikasi. Kesalahan terjadi pada kategori Transaksi, yaitu 9 data diprediksi dengan benar sebagai Transaksi, sedangkan 1 data diprediksi sebagai Saldo. Sementara itu, kategori Top Up, Saldo, Login, Verifikasi, Aplikasi, dan Layanan Pelanggan masing-masing berhasil diklasifikasikan dengan benar sebanyak 10 data. Kesalahan antara Transaksi dan Saldo menunjukkan adanya kemiripan konteks pada keluhan, misalnya transaksi yang gagal tetapi saldo pengguna ikut terpengaruh. Secara keseluruhan, confusion matrix ini menunjukkan bahwa IndoBERT Baseline memiliki kemampuan klasifikasi yang sangat baik, dengan hanya satu kesalahan prediksi dari 70 data testing atau tingkat akurasi sebesar 98,57%. Adapun hasilnya bisa dilihat pada tabel 5 berikut ini:

Tabel 5. Hasil *Confusion Matrix* IndoBERT Baseline

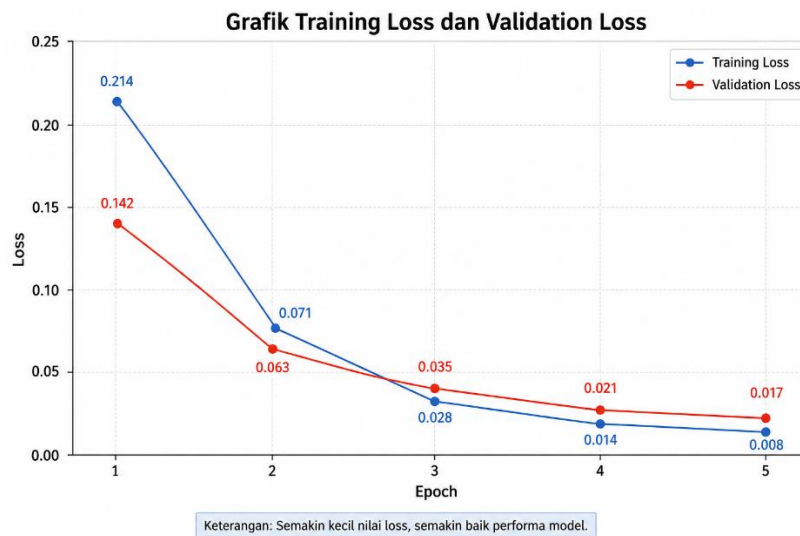
Aktual \ Prediksi	Transaksi	Top Up	Saldo	Login	Verifikasi	Aplikasi	Layanan Pelanggan
Transaksi	9	0	1	0	0	0	0
Top Up	0	10	0	0	0	0	0
Saldo	0	0	10	0	0	0	0
Login	0	0	0	10	0	0	0
Verifikasi	0	0	0	0	10	0	0
Aplikasi	0	0	0	0	0	10	0
Layanan Pelanggan	0	0	0	0	0	0	10

Berdasarkan *Confusion Matrix* IndoBERT Optimasi, seluruh 70 data testing berhasil diklasifikasikan dengan benar pada tujuh kategori, sehingga tidak terdapat kesalahan prediksi. Hasil ini menunjukkan bahwa model mampu membedakan karakteristik dan konteks setiap jenis keluhan dengan sangat baik, yang sejalan dengan nilai *accuracy*, *precision*, *recall*, dan *F1-score* sebesar 100%. Dalam konteks Manajemen Layanan TI, performa tersebut menunjukkan potensi IndoBERT untuk membantu otomatisasi pengelompokan keluhan pada *service desk* dan incident management. Namun, hasil perlu diuji kembali menggunakan data keluhan pengguna nyata karena dataset penelitian masih terbatas. Berikut hasil dari *Confusion Matrix* IndoBERT Optimasi yang bisa dilihat pada tabel 6 dibawah ini:

Tabel 6. Hasil *Confusion Matrix* IndoBERT Optimasi

Aktual \ Prediksi	Transaksi	Top Up	Saldo	Login	Verifikasi	Aplikasi	Layanan Pelanggan
Transaksi	10	0	0	0	0	0	0
Top Up	0	10	0	0	0	0	0
Saldo	0	0	10	0	0	0	0
Login	0	0	0	10	0	0	0
Verifikasi	0	0	0	0	10	0	0
Aplikasi	0	0	0	0	0	10	0
Layanan Pelanggan	0	0	0	0	0	0	10

Berikut grafik berdasarkan hasil eksperimen sementara IndoBERT Optimasi yang bisa dilihat pada gambar 2 dibawah ini:



Gambar 2 Grafik Training Loss dan Validation Loss

Grafik menunjukkan bahwa selama proses *fine-tuning* IndoBERT, *training loss* dan *validation loss* sama-sama mengalami penurunan dari *epoch* 1 hingga *epoch* 5. *Training loss* menurun dari 0,214 menjadi 0,008, sedangkan *validation loss* menurun dari 0,142 menjadi 0,017. Penurunan tersebut menunjukkan bahwa model semakin mampu mempelajari pola dan karakteristik teks keluhan pengguna *e-wallet*, sementara nilai *validation loss* juga menunjukkan perbaikan performa pada data validasi selama proses pelatihan. Pada *epoch* ke-5 diperoleh nilai *loss* terendah untuk kedua data, sehingga konfigurasi tersebut memberikan hasil pembelajaran terbaik pada eksperimen ini. Selisih antara *training loss* dan *validation loss* pada *epoch* ke-5 relatif kecil, sehingga berdasarkan kurva *loss* tidak terlihat pola divergensi yang jelas antara kedua nilai *loss* selama proses pelatihan. Namun, kondisi tersebut belum cukup untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik atau sepenuhnya terbebas dari *overfitting*. Evaluasi generalisasi perlu didukung oleh hasil pengujian pada data yang tidak digunakan selama proses pelatihan dan pemilihan hyperparameter, serta metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Oleh karena itu, kurva *loss* dalam penelitian ini digunakan sebagai indikator proses pembelajaran model, sedangkan kemampuan generalisasi model dinilai berdasarkan hasil evaluasi pada data pengujian.

KESIMPULAN

Berikan Hasil penelitian menunjukkan bahwa penggunaan IndoBERT dapat membantu mengkategorikan keluhan pengguna *e-wallet* ke dalam beberapa kategori layanan: transaksi, *top up*, saldo, *login*, verifikasi, aplikasi, dan layanan pelanggan. Model telah menunjukkan kinerja yang sangat baik pada data pengujian berkat prosedur penelitian yang mencakup pengumpulan data, *preprocessing*, pelabelan, pembagian dataset, pemodelan IndoBERT, optimasi *hyperparameter*, dan evaluasi. Hasil analisis menunjukkan bahwa IndoBERT *Baseline* memiliki ketepatan sebesar 98,57%, dan IndoBERT Optimasi memiliki ketepatan, ketepatan, *recall*, dan skor F1 sebesar 100%. Selain itu, *Confusion Matrix* IndoBERT Optimasi menunjukkan bahwa seluruh 70 data tes yang berhasil diklasifikasikan dengan benar. Hasil menunjukkan bahwa kemampuan klasifikasi dapat ditingkatkan melalui optimasi model. Ini juga dapat membantu proses pengelompokan keluhan secara otomatis dalam manajemen keluhan dan meja layanan TI. Dengan demikian, penelitian ini telah sesuai dengan tujuan yang ditetapkan pada Pendahuluan, yaitu mengoptimalkan klasifikasi keluhan pengguna *e-wallet* menggunakan IndoBERT serta mengkaji pemanfaatannya untuk mendukung pengelolaan layanan TI. Penelitian selanjutnya dapat menggunakan dataset *Google Play* yang lebih besar, menggunakan lebih dari satu anotasi, menambahkan kategori dan tingkat prioritas keluhan, dan menggunakan lebih banyak anotasi. Selain itu, model dapat dikembangkan menjadi sistem klasifikasi keluhan yang terintegrasi dengan meja layanan. Ini akan memungkinkan pengelola layanan TI untuk menggunakan hasil klasifikasi untuk evaluasi kualitas layanan TI, routing tiket, penentuan prioritas, dan identifikasi layanan yang bermasalah. Pengembangan lebih lanjut juga dapat menerapkan AI yang menjelaskan agar hasil klasifikasi lebih mudah dipahami dan dipertanggungjawabkan.

REFERENSI

- [1] D. D. Darmansah, "Perancangan Sistem Informasi Pengolahan Jadwal Mata Pelajaran Siswa Secara Online Di Smpn 31 Padang Berbasis Web," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 3, pp. 451–465, 2020, doi: 10.35957/jatisi.v7i3.490.
- [2] S. Sugiono and R. D. Nugraheni, "Cashless Society: Cluster Analysis of Electronic Payment Users in E-Commerce," *JEJAK J. Econ. Policy*, vol. 16, no. 2, pp. 286–301, 2023, doi: doi.org/10.15294/jejak.v16i2.42451.
- [3] K. Zainurrohman, D. Y. Kristiyanto, and D. Darmansah, "Evaluasi Usability Website Sistem Informasi Pelayanan Perizinan Banyumas (SIPANJIMAS) Menggunakan Metode System Usability Scale (SUS)," *J. Sist. Komput. dan Inform.*, vol. 3, no. 4, p. 510, 2022, doi: 10.30865/json.v3i4.4141.
- [4] A. Sastra and A. -, "Analisis Penggunaan Digital Payment QRIS pada Pelaku Usaha Mikro dan Kecil," *J. Ilm. Adm. Bisnis dan Inov.*, vol. 8, no. 1, pp. 15–29, 2024, doi: 10.25139/jiabi.v8i1.7867.
- [5] B. Banutama, A. J. Hatta, F. Biyanto, and D. R. Kristiana, "Keputusan Penggunaan E-Wallet Sebagai Alat Transaksi Digital," *J. Ilm. Akunt. dan Keuang.*, vol. 3, no. 4, pp. 301–318, 2025, doi: 10.24034/jiaku.v3i4.7025.
- [6] I. Alika Zuhaira and F. A. Alijoyo, "Evaluasi Proses Bisnis Dompot Digital Menggunakan Technology Acceptance Model," *J. Algoritma*, vol. 21, no. 2, pp. 64–72, 2024, doi: 10.33364/algoritma/v.21-2.1690.
- [7] E. A. Lau and U. Kulsum, "Becoming a Cashless Society: The Role of QRIS from the Z-Generation Student's Perspective," *J. Account. Strateg. Financ.*, vol. 6, no. 1, pp. 172–191, 2023, doi: 10.33005/jasf.v6i1.404.
- [8] A. Maghfur and Indrawati, "QRIS Adoption in West Java: Evidence from A Muslim-Majority Region," *J. Islam. Econ. Bus.*, vol. 5, no. 2, pp. 613–631, 2026, doi: 10.15575/jieb.v5i2.54607.
- [9] R. R. T. Jati, Saputr Restu, Dewi Winarni, "Analisis Penggunaan Aplikasi Dompot Digital di UMKMKota Depok," *J. Adm. Prof.*, vol. 04, no. 1, pp. 57–64, 2023, doi: doi.org/10.32722/jap.v4i1.6084.
- [10] W. Agustiono, Y. D. Prasetya, and Y. Kustiyahningsih, "Pengukuran Usability Aplikasi E-Wallet dengan Model PACMAD Menggunakan Metode Fuzzy-AHP dan TOPSIS," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 1, pp. 12–20, 2023, doi: 10.25077/teknosi.v9i1.2023.12-20.
- [11] A. Marsel Pratama and N. Cahyono, "Optimasi Kinerja Metode Klasifikasi Dengan Random Search Pada Analisis Sentimen Aplikasi Korlantas Polri," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4073–4080, 2025, doi: 10.36040/jati.v9i3.13480.
- [12] F. Damayanti, A. Rahim, and N. Azmi Verdikha, "Analisis Sentimen Ulasan Pengguna Terhadap Aplikasi K24Klik Di Google Play Store Dengan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3224–3230, 2025, doi: 10.36040/jati.v9i2.13298.
- [13] A. P. Mareta et al., "Perbandingan Metode Naïve Bayes, Decision Tree, Dan Knn Dalam Analisis Sentimen Aplikasi Gojek Di Playstore," *J. Sist. Inf.*, vol. 7, no. 2, pp. 725–734, 2025, doi: doi.org/10.31849/zjf8x279.
- [14] N. Nurzaman, N. Suarna, and W. Prihartono, "Analisis Sentimen Ulasan Aplikasi Threads Di Google Playstore Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 967–974, 2024, doi: 10.36040/jati.v8i1.8708.
- [15] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *Ilk. J. Ilm.*, vol. 14, no. 3, pp. 348–354, 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.
- [16] Wildan Amru Hidayat and V. R. S. Nastiti, "Perbandingan Kinerja Pre-Trained Indobert-Base Dan Indobert-Lite Pada Klasifikasi Sentimen Ulasan Tiktok Tokopedia Seller Center Dengan Model Indobert," *JSii (Jurnal Sist. Informasi)*, vol. 11, no. 2, pp. 13–20, 2024, doi: 10.30656/jsii.v11i2.9168.
- [17] S. Apriliani, A. Erfina, and C. Warman, "Fine-Tuned IndoBERT for Aspect-Based Sentiment Analysis of Indonesian Five-Star Hotel Reviews," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 437–445, 2025, doi: 10.32736/sisfokom.v14i4.2491.
- [18] Sandy Sanjaya, Rangga Gelar Guntara, and Syti Sarah Maesaroh, "Sentiment Analysis of LinkAja Digital Wallet Application Reviews on Google Play Store using Transfer Learning IndoBERT," *INOVTEK Polbeng - Seri Inform.*, vol. 10, no. 3, pp. 1730–1740, 2025, doi: 10.35314/afjx7b16.
- [19] F. Rafiandi Andhika, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe," *Metik J.*, vol. 9, p. 2025, 2025, doi: 10.47002/metik.v9i2.1098.
- [20] Y. A. Singgalen, "Performance Analysis of IndoBERT for Sentiment Classification in Indonesian Hotel Review Data," *J. Inf. Syst. Res.*, vol. 6, no. 2, pp. 976–986, 2025, doi: 10.47065/josh.v6i2.6505.

- [21] S. Fauzan, S. Sukaridhoto, and K. Khoironi, "Automatic Network Failure Detection Feature in IT Helpdesk Application PKG V2 at PT. Petrokimia Gresik," *Appl. Technol. Comput. Sci. J.*, vol. 7, no. 2, pp. 128–135, 2025, doi: 10.33086/atcsj.v7i2.7768.
- [22] Yulio Ferdinand, M. Lubis, and O. N. Pratiwi, "Classification of University IT Helpdesk Tickets Using Support Vector Machine with Hyperparameter Optimization," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 14, no. 4, pp. 517–524, 2025, doi: 10.32736/sisfokom.v14i4.2413.
- [23] R. M. Caturkusuma, F. Alzami, A. Nurhindarto, M. T. Sulistiyono, C. Irawan, and Y. Kusumawati, "Predicting IT Incident Duration using Machine Learning: A Case Study in IT Service Management," *Sinkron*, vol. 9, no. 1, pp. 8–19, 2025, doi: 10.33395/sinkron.v9i1.14310.