

SENTIMENT ANALYSIS OF LINKAJA DIGITAL WALLET APPLICATION REVIEWS ON GOOGLE PLAYSTORE USING TRANSFER LEARNING INDOBERT

ANALISIS SENTIMEN APLIKASI DOMPET DIGITAL LINKAJA DI GOOGLE PLAYSTORE MENGGUNAKAN TRANSFER LEARNING INDOBERT

Sandy Sanjaya¹, Rangga Gelar Guntara², Syti Sarah Maesaroh³

^{1,2,3}Universitas Pendidikan Indonesia,

Jl. Dadaha No.18, Kahuripan, Kec. Tawang, Kota. Tasikmalaya, Jawa Barat 46115

email sandyysanjayaa@upi.edu¹, *ranggagelar@upi.edu², sytisarah@upi.edu³

Abstract - The LinkAja digital wallet receives an average rating of 3.5 on Google Play Store despite having a higher number of user reviews than its competitors, indicating a strong need for data-driven evaluation of user satisfaction. This study performs sentiment classification on LinkAja user reviews using the IndoBERT model implemented within the CRISP-DM framework. A total of 1,483 reviews posted from January 1 to May 31, 2025 were analyzed through automatic labeling using a pretrained IndoBERT sentiment model and validated using an 80:20 hold-out scheme. Model performance was evaluated using accuracy, F1-score, and the Matthews Correlation Coefficient (MCC) to address class imbalance. The results show high classification performance with 95% accuracy, a macro F1-score of 0.92, a weighted F1-score of 0.94, and an MCC of 0.90. Sentiment distribution reveals a dominance of negative sentiments at 59.5%, followed by positive (26.1%) and neutral (14.4%) sentiments. Theoretically, this study reinforces the superiority of IndoBERT over conventional machine learning methods for Indonesian sentiment analysis. Practically, the findings provide actionable insights into service improvements, particularly regarding transaction stability and system reliability.

Keywords - Sentiment Analysis, IndoBERT, Transfer Learning, LinkAja, CRISP-DM

Abstrak - Rata-rata rating aplikasi LinkAja di Google Play Store hanya mencapai 3,5 meskipun jumlah ulasannya lebih tinggi dibandingkan dompet digital lainnya, yang mengindikasikan tingkat ketidakpuasan pengguna yang lebih besar dan memerlukan evaluasi berbasis data. Penelitian ini melakukan klasifikasi sentimen ulasan pengguna LinkAja menggunakan model IndoBERT dalam kerangka CRISP-DM. Sebanyak 1.483 ulasan periode 1 Januari–31 Mei 2025 dianalisis melalui pelabelan otomatis menggunakan model sentimen pralatih IndoBERT dan divalidasi menggunakan skema hold-out 80:20. Evaluasi model memanfaatkan akurasi, F1-score, dan *Matthews Correlation Coefficient* (MCC) untuk mengakomodasi ketidakseimbangan kelas. Hasil menunjukkan performa model yang sangat baik dengan akurasi 95%, macro F1-score 0,92, weighted F1-score 0,94, dan MCC 0,90. Distribusi sentimen memperlihatkan dominasi sentimen negatif sebesar 59,5%, disusul positif 26,1%, dan netral 14,4%. Implikasi teoretis penelitian ini menegaskan keunggulan IndoBERT dibandingkan metode konvensional dalam klasifikasi sentimen berbahasa Indonesia. Secara praktis, hasil analisis memberikan wawasan langsung mengenai aspek layanan yang perlu diprioritaskan, terutama stabilitas transaksi dan keandalan sistem.

Kata Kunci - Analisis Sentimen, IndoBERT, Transfer Learning, LinkAja, CRISP-DM

I. PENDAHULUAN

Kemajuan teknologi informasi dan komunikasi telah mengubah pola hidup masyarakat menjadi lebih efisien dan praktis [1]. Digitalisasi kini mencakup berbagai layanan, termasuk teknologi finansial (*fintech*) yang menyediakan sistem pembayaran berbasis elektronik sebagai alternatif transaksi konvensional [2]. Salah satu penyedia layanan tersebut adalah LinkAja, platform dompet digital yang dikembangkan melalui kolaborasi beberapa Badan Usaha Milik Negara (BUMN) dan beroperasi di bawah PT Fintek Karya Nusantara sejak tahun 2019.

Meskipun memiliki dukungan ekosistem yang kuat, LinkAja masih menghadapi sejumlah kendala layanan seperti kegagalan transaksi, gangguan teknis, serta penanganan keluhan yang belum optimal [2]. Data pada 1 Juni 2025 menunjukkan bahwa LinkAja memperoleh rating rata-rata 3,5 dari 785 ribu ulasan di Google Play Store, lebih rendah dibandingkan GoPay (4,4 dari 753 ribu ulasan) dan ShopeePay (4,8 dari 550 ribu ulasan) pada periode yang sama [3][4][5]. Namun demikian, jumlah ulasan yang lebih besar mengindikasikan tingginya keterlibatan pengguna dalam menyampaikan opini, sehingga perlu dilakukan evaluasi sentimen secara sistematis untuk memperoleh pemahaman yang objektif terhadap persepsi publik.

Analisis sentimen berbasis *text mining* merupakan pendekatan yang tepat karena mampu mengolah data ulasan dalam jumlah besar dan mengklasifikasikannya ke dalam sentimen tertentu secara terukur [6][7]. *Natural Language Processing* (NLP) sebagai bidang interdisipliner antara ilmu komputer, linguistik, dan kecerdasan buatan, berfokus pada pengembangan sistem yang mampu memahami bahasa manusia [8]. Dalam konteks bahasa Indonesia, Pemrosesan teks menuntut model yang mampu memahami bahasa informal dan konteks semantik. IndoBERT menjadi salah satu model pra-latih berbasis *transformer* yang banyak digunakan dalam penelitian NLP berbahasa Indonesia. Berbagai studi menunjukkan bahwa IndoBERT unggul dalam tugas klasifikasi sentimen, melampaui model multibahasa serta algoritma pembelajaran mesin konvensional seperti KNN, SVM, Naïve Bayes, dan Random Forest [9][10][11][12]. Benchmark IndoLEM turut memperkuat temuan tersebut bahwa IndoBERT memiliki performa terbaik pada berbagai tugas NLP berbahasa Indonesia, sehingga relevan digunakan untuk memahami persepsi pengguna terhadap aplikasi digital seperti LinkAja [13][14][15].

Penelitian analisis sentimen terkait ulasan pengguna LinkAja di Google Play Store masih terbatas, khususnya yang memanfaatkan IndoBERT sebagai metode utama pada konteks data yang sama. Belum terdapat penelitian yang secara empiris membandingkan performa model pra-latih berbasis *transformer* dalam mengklasifikasikan sentimen pengguna LinkAja, sehingga potensi peningkatan akurasi model belum dapat terverifikasi. Selain itu, sebagian besar penelitian sebelumnya belum menerapkan kerangka analitik yang terstruktur seperti *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang penting untuk mendukung replikasi studi dan penerapan hasil dalam kebutuhan industri [6].

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan menganalisis sentimen pengguna LinkAja pada Google Play Store menggunakan model IndoBERT dengan menerapkan kerangka CRISP-DM secara sistematis. Penelitian ini diharapkan memberikan kontribusi empiris yang lebih kuat terkait persepsi pengguna serta menjadi landasan dalam peningkatan kualitas layanan pada sektor pembayaran digital di Indonesia.

II. SIGNIFIKANSI STUDI

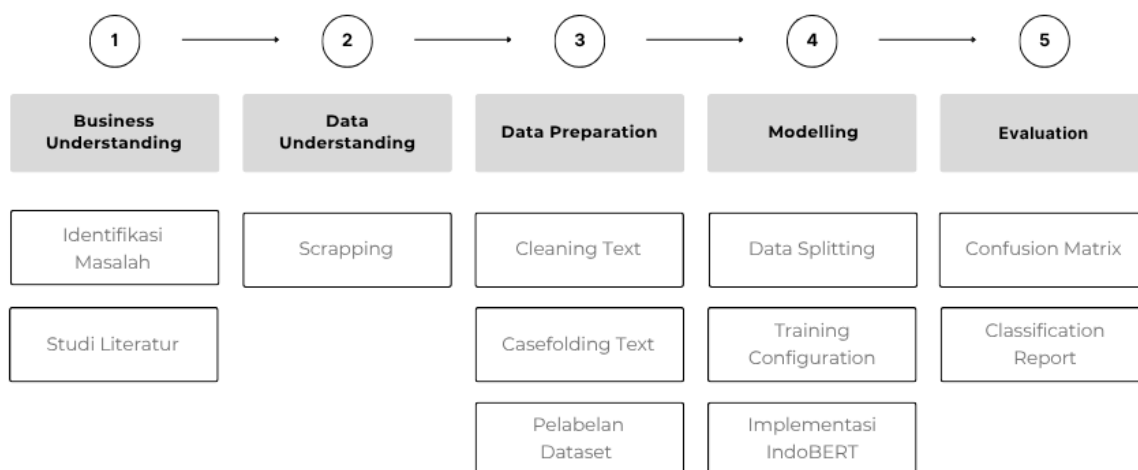
A. Penelitian Terdahulu

Penelitian terdahulu mengenai analisis sentimen aplikasi LinkAja menggunakan algoritma Naïve Bayes dan Random Forest untuk mengklasifikasikan ulasan pengguna di Google Play Store ke dalam dua kategori sentimen. Hasilnya menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 82%, melampaui Naïve Bayes yang memperoleh 79%[16]. Selain itu, penelitian lain yang menerapkan IndoBERT berbasis *deep learning* pada analisis sentimen komentar mahasiswa menunjukkan peningkatan performa dengan akurasi 85%, precision 86%, recall 80%, dan F1-score 83% [17]. Temuan tersebut menegaskan bahwa IndoBERT menghasilkan akurasi klasifikasi sentimen yang lebih unggul dibanding algoritma pembelajaran mesin konvensional.

Kontribusi teoretis penelitian ini terletak pada pembuktian empiris bahwa IndoBERT memiliki performa klasifikasi sentimen yang lebih akurat dibandingkan metode konvensional pada konteks pembayaran digital. Penelitian ini mengisi kesenjangan studi sebelumnya yang belum mengevaluasi kinerja IndoBERT langsung pada ulasan pengguna LinkAja, sehingga memperluas landasan ilmiah penggunaan model *transformer* berbahasa Indonesia dalam analisis sentimen pada sektor *fintech*. Penelitian ini juga memberikan kontribusi praktis berupa penyediaan wawasan berbasis data mengenai persepsi dan pengalaman pengguna yang dapat dimanfaatkan sebagai dasar perbaikan layanan, pengembangan fitur, serta peningkatan daya saing LinkAja secara berkelanjutan. Temuan penelitian ini diharapkan mampu menjadi rujukan dalam riset dan pengembangan model analitik serupa pada platform digital lainnya di Indonesia.

B. Metode Penelitian

Penelitian ini dilakukan secara sistematis dengan menggunakan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk menggambarkan keseluruhan proses analisis data. Kerangka kerja ini dipilih karena dinilai efektif dalam mendukung pengolahan data serta memastikan solusi berbasis data tetap selaras dengan tujuan bisnis dan organisasi, sehingga banyak digunakan dalam proyek *data mining* maupun *machine learning*. CRISP-DM sendiri terdiri atas lima tahapan utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, dan *Evaluation* [18][19][20][21]. Setiap tahapannya disusun secara runtut agar hasil yang diperoleh selaras dengan tujuan penelitian. Alur lengkap dari proses penelitian ini ditampilkan pada ilustrasi berikut:



Gambar 1. Metode Penelitian CRISP-DM\

Metode penelitian menggunakan metode CRISP-DM memiliki tahapan sebagai berikut.

Business Understanding, Tahap *Business Understanding* berfokus pada perumusan serta pemahaman tujuan dan kebutuhan utama penelitian, dengan menitikberatkan pada aspek yang relevan dengan konteks bisnis maupun domain kajian [22]. Tahap ini bertujuan mengeksplorasi sentimen pengguna terhadap aplikasi LinkAja melalui analisis ulasan di Google Play Store untuk memperoleh pemahaman komprehensif mengenai persepsi pengguna. Selain itu, tahap ini digunakan untuk menentukan pendekatan analisis sentimen yang optimal sebagai dasar penerapan metode *Transfer Learning* IndoBERT guna menghasilkan model yang lebih akurat dan berorientasi data.

Data Understanding, Tahap *Data Understanding* merupakan proses awal yang mencakup pengumpulan, eksplorasi, dan deskripsi data melalui penentuan atribut yang relevan [22]. Pada tahap ini, penelitian mengidentifikasi dan menganalisis struktur serta karakteristik data ulasan pengguna yang diperoleh melalui teknik *scraping* dari Google Play Store. Data dikumpulkan selama periode 1 Januari–31 Mei 2025 menggunakan Google Colaboratory, platform berbasis *cloud* yang memungkinkan pengguna menjalankan dan berbagi *notebook Jupyter* [23]. menghasilkan himpunan data representatif dengan atribut utama, meliputi nama pengguna, tanggal ulasan, *rating*, dan isi teks. Atribut tersebut menjadi komponen krusial dalam mendukung analisis sentimen secara komprehensif.

Data Preparation, Tahap *Data Preparation* melibatkan serangkaian proses untuk menghasilkan himpunan data siap analisis. Data mentah diolah melalui tahapan *preprocessing* guna membersihkan, menormalkan, dan menyesuaikan formatnya dengan kebutuhan model serta algoritma yang digunakan [22]. Selanjutnya, pelabelan sentimen dilakukan menggunakan *Pre-trained IndoBERT Sentiment Model* agar data yang dihasilkan memiliki kualitas dan konsistensi tinggi dalam mendukung tahap analisis berikutnya. Proses *preprocessing* mencakup beberapa tahapan utama, yaitu

- a) **Cleaning**, Tahapan awal *preprocessing* adalah *cleaning* atau pembersihan data, yang bertujuan menyederhanakan struktur teks dengan menghapus elemen tidak relevan, seperti URL, *hashtag*, *mention*, spasi berlebih, dan tanda baca. Proses ini memastikan efisiensi analisis serta konsistensi representasi data, terutama pada teks media sosial yang kerap memuat bahasa nonformal, seperti emotikon, singkatan, dan *slang* [24].
- b) **Case Folding**, *Case folding* merupakan tahap normalisasi teks dengan mengonversi seluruh huruf menjadi bentuk kecil (*lowercase*). Proses ini penting dalam analisis teks informal, seperti pada media sosial, karena kapitalisasi sering diabaikan untuk menciptakan gaya komunikasi kasual. Dengan penerapan *case folding*, sistem dapat mengenali kesetaraan kata tanpa terpengaruh perbedaan kapitalisasi, sehingga meningkatkan akurasi pemrosesan data [24].
- c) **Pelabelan Dataset**, Pelabelan dilakukan menggunakan model IndoBERT *pre-trained indonesia-bert-sentiment-classification* yang telah dilatih khusus untuk mengenali sentimen dalam Bahasa Indonesia. Model ini mengklasifikasikan ulasan ke dalam tiga label: Negatif (0), Netral (1), dan Positif (2), berdasarkan probabilitas tertinggi dari output prediksi.

Modelling, Tahap *Modeling* berfokus pada pembangunan model klasifikasi sentimen menggunakan arsitektur IndoBERT dengan *pre-trained model indobenchmark/indobert-base-p1*. Proses ini meliputi pemilihan, evaluasi, serta penyesuaian model dan algoritma agar sesuai dengan permasalahan penelitian. Kalibrasi parameter dilakukan untuk memperoleh konfigurasi optimal guna mencapai akurasi terbaik [22]. Dataset dibagi menjadi 80% data latih dan 20% data uji sesuai praktik umum pada pelatihan *deep learning* [25]. Tokenisasi diterapkan menggunakan *tokenizer* bawaan IndoBERT dengan pengaturan *padding* dan *truncation* agar panjang input seragam. Model kemudian dilatih berdasarkan konfigurasi *hyperparameter* yang disajikan pada tabel berikut.

Tabel 1. Konfigurasi Pelatihan Model IndoBERT

Parameter	Nilai	Keterangan
Epoch	10	Jumlah iterasi pelatihan penuh terhadap seluruh data latih
Batch Size	64	Ukuran <i>mini-batch</i> dalam setiap iterasi pelatihan
Learning Rate	3e-5	Laju pembelajaran awal
Weight Decay	0.01	Regulasi bobot untuk menghindari <i>overfitting</i>
Warmup Steps	800	Jumlah langkah awal tanpa penurunan <i>learning rate</i>
Tokenizer	IndoBERT Tokenizer	Tokenizer dari <i>pre-trained model</i> indobenchmark/indobert-base-p1
Padding & Truncation	Enabled	Penyesuaian panjang input agar seragam di setiap batch
Logging Steps	20	Frekuensi pencatatan log selama proses pelatihan
Evaluation Strategy	Per epoch	Evaluasi dilakukan di akhir setiap <i>epoch</i>
Model Checkpoint	Best accuracy	Menyimpan model dengan akurasi validasi tertinggi
Early Stopping	Patience = 2 epochs	Menghentikan pelatihan jika tidak ada peningkatan dalam 2 <i>epoch</i> berturut-turut

Selama pelatihan, dilakukan pencatatan log setiap 20 langkah serta pemilihan model terbaik berdasarkan akurasi validasi tertinggi pada setiap *epoch*. Untuk mencegah terjadinya *overfitting*, diterapkan pula mekanisme *EarlyStopping* dengan toleransi sebanyak dua *epoch* tanpa peningkatan performa.

Evaluatio, Tahap *Evaluation* berfungsi untuk memverifikasi kualitas model hasil *Modeling* dan menilai kesesuaiannya terhadap kriteria keberhasilan yang telah ditetapkan. Tujuan utamanya adalah mengukur kinerja model klasifikasi sentimen berbasis *transfer learning* dengan arsitektur IndoBERT. Hasil evaluasi digunakan untuk menentukan kelayakan model sebelum tahap implementasi atau kebutuhan penyesuaian guna meningkatkan performa [22]. Evaluasi dilakukan menggunakan metrik klasifikasi *accuracy*, *precision*, *recall*, dan *F1-score* yang diperoleh dari *confusion matrix*. Matriks ini menggambarkan distribusi hasil prediksi terhadap data uji dengan membandingkan label aktual dan label yang diprediksi. Struktur *confusion matrix* yang digunakan ditunjukkan sebagai berikut:

Tabel 2. Confusion Matrix

		Predicted Class		
		Negatif	Netral	Positif
Actual Class	Negatif	T _{Negatif}	F _{Netral}	F _{Positif}
	Netral	F _{Negatif}	T _{Netral}	F _{Positif}
	Positif	F _{Negatif}	F _{Netral}	T _{Positif}

Selanjutnya, nilai dari masing-masing metrik dihitung dengan rumus sebagai berikut:

- a) **Accuracy** digunakan untuk mengukur proporsi prediksi yang benar terhadap keseluruhan data uji, baik pada klasifikasi positif maupun negatif. Namun, dalam kondisi data yang tidak seimbang, metrik ini berpotensi menimbulkan interpretasi yang kurang tepat karena tidak memperhitungkan distribusi kelas [26]. Secara umum, formula akurasi dapat dituliskan sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

- b) **Precision**, atau *positive predictive value* merupakan ukuran seberapa tepat model dalam mengidentifikasi kelas positif dari seluruh prediksi positif yang dihasilkan [27]. Secara umum, *precision* dirumuskan sebagai:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

- c) **Recall**, juga dikenal sebagai sensitivitas atau *true positive rate*, menunjukkan sejauh mana model mampu mendeteksi seluruh instance aktual dari kelas positif [26]. Rumus umum *recall* adalah:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- d) **F-1 Score** merupakan rata-rata harmonik antara *precision* dan *recall*, yang banyak digunakan pada kondisi data dengan distribusi kelas yang tidak seimbang. Metrik ini berfungsi untuk memberikan keseimbangan antara kemampuan model dalam mendeteksi kelas positif sekaligus memastikan kebenaran prediksi terhadap kelas tersebut [27]

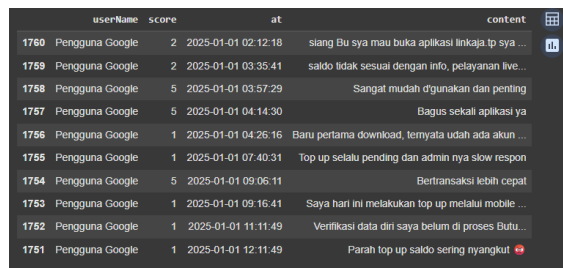
$$F1 \text{ Score} = 2 \times \frac{(Precision \times Recall)}{Precision+Recall} \quad (4)$$

III. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil yang diperoleh pada setiap tahapan penelitian yang dilaksanakan dengan mengacu pada metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Seluruh hasil disajikan secara sistematis sesuai dengan urutan tahapan yang telah ditetapkan.

Business Understanding, Data Google Play Store menunjukkan bahwa LinkAja memperoleh rating rata-rata lebih rendah dibandingkan GoPay dan ShopeePay, meskipun memiliki jumlah ulasan lebih banyak. Fenomena ini memunculkan pertanyaan terkait kecenderungan sentimen pengguna, baik positif, netral, maupun negatif. Penelitian ini menganalisis ulasan pengguna LinkAja menggunakan model *IndoBERT* berbasis *transfer learning* untuk mengidentifikasi potensi permasalahan layanan. Hasilnya diharapkan memberikan rekomendasi strategis bagi peningkatan kualitas aplikasi.

Data Understanding, Pengumpulan data diawali dengan impor *library* yang diperlukan untuk membentuk *dataset*. Data berupa ulasan pengguna aplikasi LinkAja diperoleh dari Google Play Store melalui Google Colaboratory menggunakan Python dan *library google-play-scraper*. Proses ini menghasilkan 1.483 ulasan dalam rentang 1 Januari–31 Mei 2025, yang kemudian disimpan dalam format CSV guna mempermudah tahap analisis sentimen berikutnya.



	username	score	at	content
1760	Pengguna Google	2	2025-01-01 02:12:18	siang Bu saya mau buka aplikasi linkaja tp saya ...
1759	Pengguna Google	2	2025-01-01 03:35:41	saldo tidak sesuai dengan info, pelayanan live...
1758	Pengguna Google	5	2025-01-01 03:57:29	Sangat mudah dgunakan dan penting
1757	Pengguna Google	5	2025-01-01 04:14:30	Bagus sekali aplikasi ya
1756	Pengguna Google	1	2025-01-01 04:26:16	Baru pertama download, ternyata udah ada akun ...
1755	Pengguna Google	1	2025-01-01 07:40:31	Top up selalu pending dan admin nya slow respon
1754	Pengguna Google	5	2025-01-01 09:06:11	Bertransaksi lebih cepat
1753	Pengguna Google	1	2025-01-01 09:16:41	Saya hari ini melakukan top up melalui mobile ...
1752	Pengguna Google	1	2025-01-01 11:11:49	Verifikasi data diri saya belum di proses Butu...
1751	Pengguna Google	1	2025-01-01 12:11:49	Parah top up saldo sering nyangkut

Gambar 2. Dataset Hasil *Scrapping*

Gambar 2 menyajikan cuplikan visual dari data hasil *scraping* ulasan pengguna aplikasi LinkAja di Google Play Store. Data tersebut berisi sejumlah atribut penting seperti nama pengguna, skor penilaian (*rating*), waktu unggahan ulasan, serta isi teks ulasan yang menjadi objek utama dalam proses analisis sentimen. Cuplikan ini memberikan ilustrasi awal mengenai bentuk dan isi data mentah yang telah berhasil dikumpulkan dari platform sumber.

Tabel 3. Info Dataset Hasil *Scrapping*

Column	Count	Data Type
Username	1483	Object
Score	1483	Int64
At	1483	Object
Content	1483	Object

Struktur dataset hasil *scraping* terdiri atas empat atribut utama, yaitu nama pengguna (tipe data: object) yang merepresentasikan identitas pengguna, skor penilaian (int64) sebagai nilai penilaian dalam skala 1–5, waktu unggahan ulasan (object) yang mencatat tanggal publikasi ulasan, serta isi teks ulasan (object) yang memuat teks ulasan sebagai bahan utama dalam proses analisis sentimen.

Data Preparation, Tahap *data preparation* bertujuan memastikan kesiapan data teks untuk proses pemodelan. Ulasan dari Google Play Store diproses melalui tahapan *preprocessing* untuk menghapus elemen tidak relevan, menstandarkan format, dan menambahkan label sentimen guna mendukung analisis lanjutan. Prosedur ini meliputi *cleaning*, *case folding*, serta pelabelan sentimen otomatis menggunakan model *IndoBERT* pralatih. Setiap tahap dirancang untuk meningkatkan kualitas representasi data agar model dapat belajar secara optimal dan menghasilkan prediksi sentimen yang akurat.

- a) **Cleaning**, Pada tahap *cleaning*, dataset dibersihkan dari berbagai elemen yang tidak diperlukan, seperti emoji, simbol, *mention*, *hashtag*, *link* serta komponen lain yang berpotensi mengganggu akurasi dalam proses analisis sentimen.

Tabel 4. Tahap *Cleaning*

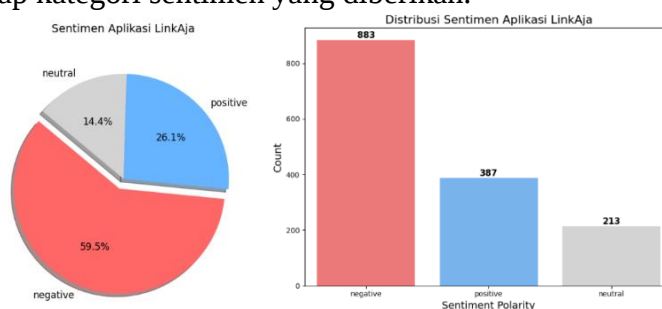
Data Ulasan	Hasil <i>Cleaning</i>
Aplikasi 'keren' yg sulit chat sama agen	Aplikasi keren yg sulit chat sama agen

- b) **Case Folding**, Seluruh teks pada data ulasan dikonversi ke dalam format huruf kecil (*lowercase*) pada tahap ini, dengan tujuan menjaga konsistensi penulisan serta mempermudah proses analisis data pada tahap berikutnya.

Tabel 5. Tahap *Case Folding*

Data Ulasan	Hasil <i>Case Folding</i>
Aplikasi 'keren' yg sulit chat sama agen	aplikasi keren yg sulit chat sama agen

- c) **Pelabelan Dataset**, Pelabelan sentimen dilakukan secara otomatis menggunakan model IndoBERT pralatih yang menghasilkan distribusi probabilitas untuk setiap kelas sentimen. Kategori dengan nilai probabilitas tertinggi ditetapkan sebagai label pada tiga kelas utama, yaitu positif, netral, dan negatif. Untuk memastikan kualitas label serta mengurangi potensi bias pelabelan otomatis, seluruh hasil pelabelan diverifikasi secara manual dengan mencocokkan konteks ulasan terhadap kategori sentimen yang diberikan.



Gambar 3. Distribusi Sentimen Aplikasi LinkAja

Berdasarkan hasil akhir pelabelan dan verifikasi, proses pelabelan pada dataset yang telah divisualisasikan memperlihatkan bahwa dari total 1.483 ulasan, sebanyak 883 ulasan (59,5%) dikategorikan sebagai sentimen negatif, 387 ulasan (26,1%) sebagai sentimen positif, dan 213 ulasan (14,4%) sebagai sentimen netral.

Modelling, Tahap pelatihan model klasifikasi sentimen dilakukan dengan pendekatan *transfer learning* menggunakan arsitektur IndoBERT. Dataset hasil praproses dan pelabelan otomatis selanjutnya dibagi menjadi 80% data latih dan 20% data uji untuk memastikan generalisasi model. Performa pelatihan dievaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, MCC digunakan sebagai metrik tambahan karena lebih mampu memberikan gambaran kinerja model secara objektif pada kondisi distribusi kelas yang tidak seimbang, termasuk pada skenario *multi-class*. Penggunaan MCC sangat dianjurkan dalam konteks klasifikasi dengan dominasi kelas tertentu, karena metrik ini mempertimbangkan seluruh elemen dalam *confusion matrix*, sehingga dapat mengurangi bias evaluasi [28].

Tabel 6. Hasil Pelatihan Model

Epoch	Validation Loss	Accuracy	F1-Score	MCC
1	0.970815	0.585859	0.440327	-0.052542
2	0.791776	0.760943	0.698887	0.553570
3	0.641613	0.781145	0.719288	0.591134
4	0.541621	0.794613	0.723026	0.623432
5	0.448706	0.814815	0.756680	0.665322
6	0.342990	0.882155	0.863113	0.788003
7	0.268533	0.921552	0.892855	0.841809
8	0.218103	0.939934	0.934282	0.890465
9	0.191918	0.929293	0.930480	0.876158
10	0.183416	0.946128	0.943837	0.904094

Hasil pelatihan model IndoBERT dengan arsitektur indobenchmark/indobert-base-p1 pada Tabel 6 menunjukkan peningkatan performa yang konsisten. Akurasi validasi meningkat dari 58,59% menjadi 94,61%, sejalan dengan penurunan *validation loss* dari 0,97 menjadi 0,18 yang menandakan proses pembelajaran berlangsung efektif dan tidak menunjukkan gejala *overfitting* secara signifikan. Walaupun akurasi sempat mendekati 100%, capaian akurasi sempurna dalam tugas NLP tidak realistis dan justru berpotensi menimbulkan *overfitting* yang dapat menurunkan kinerja pada data uji [29]. Selain itu, F1-score juga meningkat hingga mencapai 94,38%, menunjukkan keseimbangan performa pada presisi dan recall. Peningkatan *Matthews Correlation Coefficient* (MCC) dari -0,05 menjadi 0,90 semakin menguatkan reliabilitas hasil evaluasi pada distribusi kelas yang tidak seimbang, karena MCC mempertimbangkan seluruh elemen dalam *confusion matrix* sehingga mampu memberikan penilaian kinerja yang lebih objektif. Penerapan *early stopping* memastikan model terbaik tersimpan dengan tetap menjaga kemampuan generalisasi terhadap data yang belum pernah dilihat.

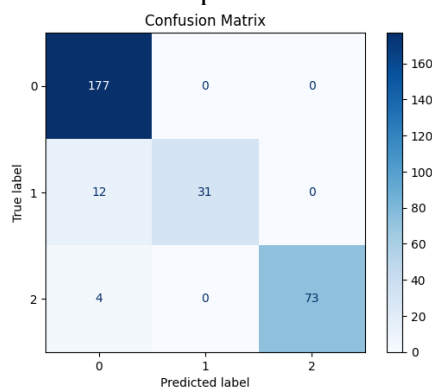
Evaluation Tahap evaluasi dilakukan untuk menilai kinerja model dalam mengklasifikasikan polaritas sentimen ulasan pengguna aplikasi LinkAja. Penilaian dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* sebagaimana ditunjukkan pada Tabel 7.

Tabel 7. *Classification Report*

Label	Precision	Recall	F1-Score	Support
0	0.92	1.00	0.96	177
1	1.00	0.72	0.84	43
2	1.00	0.95	0.97	77
Accuracy			0.95	297
Macro Avg	0.97	0.89	0.92	297
Weighted Avg	0.95	0.95	0.94	297

Matthews Correlation Coefficient (MCC): 0.9041

Hasil evaluasi menunjukkan bahwa model IndoBERT memiliki akurasi sebesar 95%, yang menandakan kemampuan tinggi dalam memprediksi sentimen secara tepat pada sebagian besar data uji. Nilai *Matthews Correlation Coefficient* (MCC) sebesar 0,9041 mengindikasikan adanya korelasi yang sangat kuat antara hasil prediksi model dan label sebenarnya, sehingga model dapat dikatakan memiliki keseimbangan performa yang baik di seluruh kelas. Selain itu, *macro average F1-score* sebesar 0,92 dan *weighted average F1-score* sebesar 0,94 menunjukkan bahwa model mampu menjaga konsistensi kinerja antar kelas, baik secara rata-rata sederhana maupun dengan mempertimbangkan proporsi jumlah data pada masing-masing kelas. Secara keseluruhan, hasil ini menegaskan bahwa model IndoBERT memiliki performa klasifikasi yang stabil dan andal.



Gambar 4. *Confusion Matrix*

Hasil *confusion matrix* menunjukkan bahwa model IndoBERT mampu mengklasifikasikan polaritas sentimen dengan akurasi tinggi sebesar 95% dari 297 data uji. Sebanyak 281 ulasan diprediksi dengan benar, terdiri atas 177 negatif, 31 netral, dan 73 positif. Kesalahan klasifikasi tergolong kecil, dengan sebagian data netral dan positif terdeteksi sebagai negatif. Dominasi pola diagonal menegaskan

kemampuan model dalam membedakan setiap kelas secara konsisten dan presisi, sehingga membuktikan efektivitas IndoBERT dalam analisis sentimen multikategori berbahasa Indonesia.

Tabel 8. Perbandingan Performa Model

Author/year	Source	Method	Accuracy	MCC
Kaeren dan Andrianingsih (2025)	Google Playstore	Random Forest	82%	
		Naïve Bayes	79%	
Baihaqi dan Munandar (2023)	Sumber Primer (Kuesioner)			
	Sumber Sekunder (e-Commerce platform)	IndoBERT	85%	
Our Study	Google Playstore	IndoBERT	95%	0,90
		Random Forest	80%	0,56
		Naïve Bayes	76%	0,63

Dibandingkan dengan penelitian sebelumnya, hasil penelitian ini menunjukkan peningkatan signifikan pada seluruh metrik evaluasi. Studi terdahulu menunjukkan bahwa algoritma konvensional seperti Random Forest dan Naïve Bayes masing-masing hanya mencapai akurasi 82% dan 79%, sementara penelitian berbasis IndoBERT sebelumnya mencatat akurasi 85%. Pada penelitian ini, IndoBERT menghasilkan peningkatan yang signifikan dengan akurasi 94,2% dan MCC 0,90, sekaligus unggul atas *baseline* pada dataset yang sama, yaitu Random Forest (80%; MCC 0,56) dan Naïve Bayes (76%; MCC 0,63). Hasil tersebut menegaskan bahwa pendekatan transfer learning IndoBERT memiliki kemampuan representasi semantik yang lebih kuat serta memberikan evaluasi yang lebih objektif pada distribusi kelas yang tidak seimbang dalam analisis sentimen aplikasi *fintech*.

Interpretasi Hasil Prediksi, Pemodelan topik menggunakan Latent Dirichlet Allocation (LDA) menunjukkan bahwa setiap kategori sentimen memiliki tiga topik utama yang merepresentasikan dimensi persepsi pengguna terhadap aplikasi LinkAja. Pada **sentimen positif**, tiga aspek utama dengan proporsi 31,98%, 31,39%, dan 36,63% menggambarkan persepsi pengguna terhadap kemudahan penggunaan (*usability*), kecepatan transaksi, serta kepuasan emosional terhadap layanan. Sementara itu, **sentimen netral** dengan proporsi 36,95%, 26,06%, dan 36,99% mencerminkan aktivitas informatif seperti autentikasi akun, proses transaksi, dan *top-up* saldo tanpa ekspresi emosional yang signifikan. Adapun **sentimen negatif**, yang memiliki proporsi 23,82%, 36,26%, dan 39,92%, didominasi oleh keluhan mengenai kegagalan login dan verifikasi, transaksi tertunda, serta ketidakstabilan sistem aplikasi yang membentuk kluster “gangguan operasional” sebagai sumber utama persepsi negatif pengguna.

Temuan tersebut memperlihatkan bahwa keandalan sistem transaksi merupakan aspek paling krusial yang perlu diprioritaskan oleh pengembang LinkAja, sejalan dengan penelitian terdahulu yang menyatakan bahwa stabilitas transaksi merupakan determinan utama kepuasan dan loyalitas pengguna layanan pembayaran digital [1]. Penerapan analisis sentimen otomatis secara berkelanjutan juga berpotensi menjadi mekanisme deteksi dini terhadap lonjakan keluhan teknis, sehingga respons terhadap gangguan layanan dapat dilakukan lebih cepat dan terukur [6]. Selain fokus pada aspek teknis, perbaikan efektivitas penanganan keluhan pengguna tetap relevan sebagai strategi pendukung peningkatan kualitas layanan, karena dimensi *reliability* dan *responsiveness* berkontribusi terhadap pembentukan persepsi layanan dan kepuasan pengguna [18].

IV. KESIMPULAN

Penelitian ini menunjukkan bahwa IndoBERT memiliki performa unggul dalam klasifikasi sentimen ulasan pengguna LinkAja. Model mencapai akurasi 95%, macro F1-score 0,92, weighted F1-score 0,94, serta MCC 0,90, yang menegaskan kinerja model tetap konsisten pada distribusi kelas yang tidak seimbang. Distribusi sentimen memperlihatkan bahwa 59,5% ulasan pengguna bersifat negatif, 26,1% positif, dan 14,4% netral, dengan mayoritas keluhan berfokus pada ketidakstabilan transaksi, kendala login, dan gangguan operasional aplikasi. Temuan ini menegaskan bahwa peningkatan stabilitas transaksi serta aspek *reliability* dan *responsiveness* perlu menjadi fokus utama dalam pengembangan layanan LinkAja. Secara teoretis, penelitian ini memperkuat bukti empiris mengenai IndoBERT lebih unggul dibandingkan algoritma pembelajaran mesin konvensional dalam tugas analisis sentimen berbahasa Indonesia. Secara praktis, hasil penelitian menyediakan wawasan berbasis data yang dapat dimanfaatkan untuk pemetaan persepsi pengguna dan penentuan fokus perbaikan layanan pembayaran digital. Penelitian ini masih memiliki keterbatasan pada aspek validasi label karena pelabelan sentimen dilakukan secara otomatis tanpa verifikasi menyeluruh berbasis anotator ahli sebagai *gold standard*. Selain itu, cakupan data hanya terbatas pada ulasan LinkAja di Google Play Store pada satu rentang waktu tertentu. Penelitian berikutnya perlu memperluas sumber data, menghadirkan proses anotasi manual yang lebih komprehensif, serta menguji performa model pada konteks operasional yang lebih dinamis agar hasil penelitian memiliki generalisasi yang lebih kuat.

REFERENSI

- [1] D. E. Oktania and T. A. Indarwati, "Pengaruh Perceived Usefulness, Perceived Ease of Use, dan Compatibility With Lifestyle Terhadap Niat Beli Di Social Commerce," *J. Ilmu Manaj. Fak. Ekon. dan Bisnis Univ. Negeri Surabaya*, vol. 10, no. 1, pp. 255–267, 2022, [Online]. Available: <https://journal.inspirasi.or.id/nomicpedia/article/view/138/53>
- [2] P. O. Yudiantara and I. G. J. A. Widagda, "The role of trust in mediating the effect of perceived usefulness and perceived ease of use on decisions to use the LinkAja digital wallet : Study on LinkAja application users in Denpasar City," *Int. J. Health Sci. (Qassim)*, vol. 6, no. July, pp. 6310–6327, 2022, [Online]. Available: <https://doi.org/10.53730/ijhs.v6nS4.11176>
- [3] Google LLC, "LinkAja – Apps on Google Play," LinkAja – Apps on Google Play. Accessed: Jun. 01, 2025. [Online]. Available: <https://play.google.com/store/apps/details?id=com.telkom.mwallet&hl=id>
- [4] Google LLC, "GoPay – Apps on Google Play," GoPay – Apps on Google Play. Accessed: Jun. 01, 2025. [Online]. Available: <https://play.google.com/store/apps/details?id=com.gojek.gopay&hl=id>
- [5] Google LLC, "ShopeePay – Apps on Google Play," ShopeePay – Apps on Google Play. Accessed: Jun. 01, 2025. [Online]. Available: <https://play.google.com/store/apps/details?id=com.shopeepay.id&hl=id>
- [6] M. R. Amalsyah, D. Kurniawan, A. Rifai, and P. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Fintech menggunakan Framework CRISP - DM dalam Penentuan Prioritas Pengembangan Produk Sentiment Analysis of Fintech Application User Reviews Using the CRISP-DM Framework for Product Development Prioritization," vol. 14, pp. 813–825, 2025.
- [7] F. Aftab *et al.*, "A Comprehensive Survey on Sentiment Analysis Techniques," *Int. J. Technol.*, vol. 14, no. 6, pp. 1288–1298, 2023, doi: 10.14716/ijtech.v14i6.6632.
- [8] V. R. Prasetyo, N. Benarkah, and V. J. Chrisintha, "Implementasi Natural Language Processing Dalam Pembuatan Chatbot Pada Program Information Technology Universitas Surabaya," *Teknika*, vol. 10, no. 2, pp. 114–121, 2021, doi: 10.34148/teknika.v10i2.370.
- [9] 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.
- [10] S. Cahyawijaya *et al.*, "IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language Generation," *EMNLP 2021 - 2021 Conf. Empir. Methods Nat. Lang. Process. Proc.*, pp. 8875–8898, 2021, doi: 10.18653/v1/2021.emnlp-main.699.
- [11] K. S. Nugroho, A. Y. Sukmadewa, H. Wuswilahaken Dw, F. A. Bachtiar, and N. Yudistira, "BERT Fine-Tuning for Sentiment Analysis on Indonesian Mobile Apps Reviews," *ACM Int. Conf.*

- Proceeding Ser.*, pp. 258–264, 2021, doi: 10.1145/3479645.3479679.
- [12] H. Imaduddin, F. Y. A'la, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 113–117, 2023, doi: 10.14569/IJACSA.2023.0140813.
 - [13] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th Int. Conf. Comput. Linguist. Proc. Conf.*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.
 - [14] Y. Asri, D. Kuswardani, W. N. Suliyanti, Y. O. Manullang, and A. R. Ansyari, "Sentiment analysis based on Indonesian language lexicon and IndoBERT on user reviews PLN mobile application," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 38, no. 1, p. 677, 2025, doi: 10.11591/ijeecs.v38.i1.pp677-688.
 - [15] I. R. Hidayat and W. Maharani, "General Depression Detection Analysis Using IndoBERT Method," *Int. J. Inf. Commun. Technol.*, vol. 8, no. 1, pp. 41–51, 2022, doi: 10.21108/ijoict.v8i1.634.
 - [16] K. Kaeren and A. Andrianingsih, "Analisis Sentimen Aplikasi LinkAja di Google Play Store Menggunakan Algoritma Naïve Bayes dan Random Forest," *J. Ris. dan Apl. Mhs. Inform.*, vol. 6, no. 02, pp. 438–447, 2025, doi: 10.30998/jrami.v6i02.13821.
 - [17] W. M. Baihaqi and A. Munandar, "Sentiment Analysis of Student Comment on the College Performance Evaluation Questionnaire Using Naïve Bayes and IndoBERT," *JUITA J. Inform.*, vol. 11, no. 2, p. 213, 2023, doi: 10.30595/juita.v11i2.17336.
 - [18] Y. A. Singgalen, "Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1870–1882, 2023, doi: 10.47065/bits.v4i4.3199.
 - [19] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.
 - [20] L. Z.S. Sudar, J. L. Imbenay, I. Budi, A. Ramadiah, P. K. Putra, and A. B. Santoso, "Textual Analysis for Public Sentiment Toward National Police Using CRISP-DM Framework," *Rev. d'Intelligence Artif.*, vol. 38, no. 1, pp. 63–72, 2024, doi: 10.18280/ria.380107.
 - [21] A. N. Ma'aly, D. Pramesti, A. D. Fathurahman, and H. Fakhurroja, "Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews Using Multi-Label Classification with a Deep Learning Algorithm," *Inf.*, vol. 15, no. 11, pp. 1–33, 2024, doi: 10.3390/info15110705.
 - [22] R. A. Casonatto, T. D. P. G. Souza, and A. M. Mariano, "Quality and Risk Management in Data Mining: A CRISP-DM Perspective," *Procedia Comput. Sci.*, vol. 242, pp. 161–168, 2024, doi: 10.1016/j.procs.2024.08.257.
 - [23] R. Gelar Guntara, "Deteksi Atap Bangunan Berbasis Citra Udara Menggunakan Google Colab dan Algoritma Deep Learning YOLOv7," *J. Manaj. Sist. Inf.*, vol. 2, no. 1, pp. 9–18, 2023, doi: 10.59431/jmasif.v2i1.156.
 - [24] M. A. Palomino and F. Aider, "Evaluating-the-Effectiveness-of-Text-PreProcessing-in-Sentiment-AnalysisApplied-Sciences-Switzerland.pdf," *Mdpi*, vol. 12, p. 8765, 2022.
 - [25] A. Rácz, D. Bajusz, and K. Héberger, "Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification," *Molecules*, vol. 26, no. 4, pp. 1–16, 2021, doi: 10.3390/molecules26041111.
 - [26] G. M. Foody, "Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient," *PLoS One*, vol. 18, no. 10 October, pp. 1–27, 2023, doi: 10.1371/journal.pone.0291908.
 - [27] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," pp. 37–63, 2020, [Online]. Available: <http://arxiv.org/abs/2010.16061>
 - [28] J. Tamura, Y. Itaya, K. Hayashi, and K. Yamamoto, "Statistical Inference of the Matthews Correlation Coefficient for Multiclass Classification," pp. 1–19, 2025, [Online]. Available: <http://arxiv.org/abs/2503.06450>
 - [29] F. Panjaitan, W. Ce, H. Oktafiandi, G. Kanugrahan, Y. Ramdhani, and V. H. C. Putra, "Evaluation of Machine Learning Models for Sentiment Analysis in the South Sumatra Governor Election Using Data Balancing Techniques," *J. Inf. Syst. Informatics*, vol. 7, no. 1, pp. 461–478, 2025, doi: 10.51519/journalisi.v7i1.1019.