Skema Pendanaan: MADYA

LAPORAN PENELITIAN

PERBANDINGAN KINERJA MODEL MACHINE LEARNING DAN INDOBERT PADA ANALISIS SENTIMEN ULASAN E-COMMERCE DENGAN DAN TANPA NORMALISASI BAHASA GAUL

TIM PENELITI

Ketua : Putri Hayati, S.ST., M.Kom. (170067)

Anggota : Abdul Muis Sobri, S.Ag., M.Kom. (140062)

FAKULTAS/PUSAT STUDI

UNIVERSITAS BUDI LUHUR

Juli 2026

HALAMAN PENGESAHAN
LAPORAN PENELITIAN

Judul Penelitian :Perbandingan Kinerja Model Machine Learning dan IndoBERT pada Analisis Sentimen Ulasan E-Commerce dengan dan tanpa Normalisasi Bahasa Gaul

Bidang Penelitian : Natural Language Processing (NLP)

Ketua Peneliti

a. Nama Lengkap : Putri Hayati, S.ST., M.Kom
b. NIP/NIDN/ID-SINTA : 170067/0308029101/6145520
c. Jabatan Fungsional : Asisten Ahli
d. Program Studi : Teknik Informatika
e. Nomor HP : 082362361515
f. Alamat e-mail : putri.hayati@budiluhur.ac.id

Anggota Peneliti (1)

a. Nama Lengkap : Abdul Muis Sobri, S.Ag., M.Kom.
b. NIP/NIDN/ID-SINTA : 140062/0324107203

Mahasiswa (1)

a. Nama Lengkap : Muhammad Rifqi Fauzan
b. NIM : 2311500553

Lama Penelitian : 6 bulan

Biaya Penelitian

a. Sumber Universitas Budi Luhur : Rp. 9.000.000,-
b. Sumber lain (sebutkan jika ada) : Rp. -

Jakarta, 30 Juli 2026

Mengetahui,
Dekan/Kepala Pusat Studi

Ketua Pelaksana

(Dr. Ir. Achmad Solichin, S.Kom., M.T.I.)
NIP 050023

(Putri Hayati, S.ST., M.Kom.)
NIP 170067

Menyetujui,
Direktur Riset dan Pengabdian Kepada Masyarakat

(Prof. Dr. Ir. Prudensius Maring, M.A.)
NIP 190043

No. Registrasi : 0 1 6 0 1 LPJ 0 7 2 6
Tanggal : 3 0 0 7 2 6 Paraf :

RINGKASAN

Pertumbuhan *e-commerce* di Indonesia telah mendorong peningkatan signifikan dalam penggunaan bahasa tidak baku atau bahasa gaul, seperti singkatan, slang, dan kata-kata tidak standar dalam ulasan produk. Kondisi ini menjadi tantangan serius dalam analisis sentimen karena memengaruhi kualitas representasi teks dan menurunkan performa model klasifikasi. Normalisasi teks umumnya digunakan dalam pemrosesan bahasa alami (NLP), namun belum banyak penelitian yang mengevaluasi pengaruh sistematis normalisasi teks terhadap performa model, khususnya pada ulasan *e-commerce* berbahasa Indonesia. Penelitian ini bertujuan mengevaluasi pengaruh normalisasi teks meliputi singkatan, bahasa gaul, dan pengulangan huruf sebagai variabel independen terhadap performa model machine learning dan IndoBERT dalam analisis sentimen. Perbandingan dilakukan melalui dua skenario eksperimen: Skenario A (tanpa normalisasi) dan Skenario B (dengan normalisasi menggunakan kamus dictionary-based). Metode yang digunakan adalah pendekatan kuantitatif dengan eksperimen komparatif. Dataset terdiri dari 10.250 ulasan dari Shopee dan Tokopedia (10.000 setelah cleaning), dilabeli secara manual oleh dua annotator dengan nilai reliabilitas Cohen's Kappa = 0,78. Distribusi sentimen: 52% positif, 31% negatif, 17% netral. Normalisasi menggunakan kamus semi-manual dengan 520 pasangan kata yang telah divalidasi ahli bahasa. Model yang dibandingkan: Naive Bayes (MultinomialNB), Support Vector Machine (kernel linear, C=1), dan IndoBERT (fine-tuning: lr=2e-5, epoch=3, batch size=16). Evaluasi menggunakan akurasi, presisi, recall, dan F1-score, serta uji paired t-test (α=0,05). Hasil menunjukkan normalisasi teks secara signifikan meningkatkan performa ketiga model ($p < 0,05$), dengan peningkatan F1-score terbesar pada Naive Bayes (+5,48%), SVM (+4,53%), dan IndoBERT (+2,82%). IndoBERT tetap menjadi model terbaik pada kedua skenario. Error analysis mengungkap singkatan umum sebagai kategori kata tidak baku paling mengganggu model konvensional, sementara code-switching menjadi tantangan utama IndoBERT.

PRAKATA

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa atas terselesaikannya laporan akhir penelitian berjudul "Perbandingan Kinerja Model Machine Learning dan IndoBERT pada Analisis Sentimen Ulasan *E-commerce* dengan dan tanpa Normalisasi Bahasa Gaul". Penelitian ini merupakan bagian dari Tahun ke-3 peta jalan penelitian lima tahun di bidang analisis sentimen teks berbahasa Indonesia, dan telah dilaksanakan sesuai rencana selama 6 bulan.

Kami menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agus Setyo Budi, M.Sc. selaku Rektor Universitas Budi Luhur.
2. Bapak Dr. Ir. Achmad Solichin, S.Kom., M.T.I. selaku Dekan Fakultas Teknologi Informasi Universitas Budi Luhur atas dukungan administratif selama pelaksanaan penelitian.
3. Prof. Dr. Ir. Prudensius Maring, M.A. selaku Direktur Riset dan Pengabdian Kepada Masyarakat Universitas Budi Luhur atas dukungan pendanaan dan fasilitas penelitian.
4. Ahli bahasa yang telah meluangkan waktu untuk memvalidasi kamus normalisasi bahasa gaul.
5. Para annotator yang telah melakukan pelabelan data ulasan secara manual dengan penuh dedikasi dan konsistensi.
6. Seluruh civitas akademika Universitas Budi Luhur yang telah memberikan dukungan dan masukan selama penelitian berlangsung.

Kami menyadari laporan ini masih memiliki keterbatasan. Kritik dan saran yang membangun sangat kami harapkan untuk penyempurnaan penelitian selanjutnya. Semoga hasil penelitian ini dapat memberikan kontribusi nyata bagi pengembangan ilmu pengetahuan, khususnya di bidang pemrosesan bahasa alami berbahasa Indonesia.

Jakarta, Juli 2026

Tim Peneliti

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang dan Rumusan Masalah

Perkembangan *e-commerce* di Indonesia meningkat signifikan dalam beberapa tahun terakhir. Nilai transaksi *e-commerce* nasional terus bertumbuh pesat, menghasilkan jutaan ulasan pengguna yang mengandung opini berharga terkait persepsi konsumen terhadap produk dan layanan. Analisis sentimen menjadi pendekatan penting dalam mengolah data teks tersebut menjadi informasi yang bernilai bagi pelaku usaha maupun penelitian (1).

Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang mengklasifikasikan teks ke dalam sentimen positif, negatif, atau netral. Berbagai metode telah dikembangkan, mulai dari *machine learning* konvensional seperti Naive Bayes dan Support Vector Machine (SVM) hingga deep learning berbasis transformer seperti IndoBERT yang memiliki kemampuan tinggi dalam memahami konteks bahasa Indonesia(2,3). Penelitian terbaru menunjukkan IndoBERT memberikan akurasi lebih baik dibandingkan metode tradisional pada berbagai dataset teks berbahasa Indonesia(4,5).

Tantangan utama pada analisis sentimen ulasan *e-commerce* adalah penggunaan bahasa tidak baku dan bahasa gaul yang masif. Pengguna kerap menggunakan singkatan seperti "gk" (tidak), variasi kata seperti "bgt" (banget), ekspresi slang seperti "mantull" (mantap betul), serta pengulangan huruf seperti "baguuus". Penggunaan bahasa informal ini menurunkan kualitas representasi teks dan berdampak negatif pada kemampuan model klasifikasi.

Normalisasi teks sering digunakan untuk mengatasi masalah ini dengan mengubah kata tidak baku menjadi bentuk baku sesuai KBBI. Namun, sebagian besar penelitian sebelumnya hanya menempatkan normalisasi sebagai tahap *preprocessing* standar tanpa mengevaluasi pengaruhnya secara sistematis (6,7). Akibatnya, belum diketahui secara jelas sejauh mana normalisasi teks berperan dalam meningkatkan performa model analisis sentimen.

Berdasarkan permasalahan tersebut, rumusan masalah penelitian ini adalah:

a. Bagaimana pengaruh normalisasi teks terhadap performa model analisis sentimen pada ulasan *e-commerce* berbahasa Indonesia?
b. Bagaimana perbandingan performa model machine learning (Naive Bayes, SVM) dan IndoBERT pada data sebelum dan sesudah normalisasi?
c. Sejauh mana masing-masing model mampu menangani variasi bahasa tidak baku tanpa proses normalisasi?

Tujuan penelitian yang ingin dicapai:

a. Mengukur pengaruh normalisasi teks terhadap performa (akurasi, F1-score) model Naive Bayes, SVM, dan IndoBERT.
b. Membandingkan peningkatan performa antara model machine learning dan IndoBERT setelah normalisasi teks.
c. Mengidentifikasi ketahanan masing-masing model terhadap bahasa tidak baku melalui error analysis pada skenario tanpa normalisasi.

1.2 Pendekatan Pemecahan Masalah

Penelitian ini menggunakan pendekatan eksperimen komparatif dengan dua skenario utama: tanpa normalisasi (Skenario A) sebagai *baseline* dan dengan normalisasi (Skenario B). Normalisasi teks diposisikan sebagai variabel independen yang diuji pengaruhnya secara sistematis. Pendekatan pemecahan masalah dilakukan melalui langkah-langkah berikut:

a. Membangun dataset ulasan *e-commerce* berlabel minimal 10.000 ulasan dari Shopee dan Tokopedia melalui web *scraping*.
b. Menyusun kamus normalisasi bahasa gaul berbasis data (target minimal 500 pasangan kata) yang divalidasi oleh ahli bahasa.
c. Menerapkan *preprocessing* standar (*case folding*, tokenisasi, *stopword* removal menggunakan Sastrawi) pada kedua skenario.
d. Melatih dan mengevaluasi model Naive Bayes, SVM, dan IndoBERT pada kedua skenario.
e. Menganalisis perbedaan performa menggunakan uji statistik paired t-test (α=0,05) untuk melihat signifikansi pengaruh normalisasi.
f. Melakukan error analysis untuk mengidentifikasi jenis kata tidak baku yang paling sulit dikenali masing-masing model.

1.3 *State of the Art* dan Kebaruan

Penelitian terbaru menunjukkan bahwa model berbasis *deep learning* seperti IndoBERT memberikan akurasi lebih baik dibandingkan metode tradisional pada berbagai *dataset* teks berbahasa Indonesia (4,5). Namun, penelitian-penelitian tersebut umumnya menempatkan normalisasi hanya sebagai tahap *preprocessing* standar tanpa menguji pengaruhnya secara terpisah sebagai variabel independen. Perbandingan posisi penelitian ini disajikan pada Tabel 1.1.

Tabel 1.1. Perbandingan dengan Penelitian Sebelumnya

Peneliti (Tahun)	Metode	Normalisasi?	Menguji Pengaruh Normalisasi ?	Dataset
Geni dkk. (2023)	IndoBERT vs ML	Ya (standar)	Tidak	Media sosial
Oswari dkk. (2024)	IndoBERT	Ya (standar)	Tidak	Ulasan aplikasi
Setiawan dkk. (2025)	BERT vs SVM	Tidak	Tidak	Ulasan *e-commerce*
Sanjaya dkk. (2025)	IndoBERT	Ya (standar)	Tidak	Berita online
Penelitian ini	**NB, SVM, IndoBERT**	**Dictionary-based**	**Ya (dua skenario)**	**Ulasan e-commerce**

Kebaruan penelitian ini mencakup empat aspek utama: (1) normalisasi teks dievaluasi sebagai variabel independen, bukan sekadar *preprocessing*; (2) perbandingan

performa dilakukan pada kondisi data yang berbeda secara eksplisit; (3) pengujian ketahanan model terhadap bahasa gaul secara sistematis melalui *error analysis*; dan (4) penyediaan kamus normalisasi bahasa gaul khusus domain *e-commerce* berbahasa Indonesia.

1.4 Peta Jalan (Road Map) Penelitian 5 Tahun

Penelitian ini merupakan bagian dari peta jalan penelitian lima tahun dalam pengembangan sistem analisis sentimen teks berbahasa Indonesia pada ulasan *e-commerce* , sebagaimana disajikan pada Gambar 1.1 berikut.

Gambar 1.1. Peta Jalan Penelitian 5 Tahun

Tahun 1-2 (Baseline Performa)

Studi komparatif algoritma klasifikasi tradisional NB, SVM, dan KNN pada teks mentah berbahasa Indonesia untuk memperoleh baseline performa model.

Tahun 3 Penelitian Saat Ini (Evaluasi Normalisasi Teks)

Evaluasi pengaruh normalisasi teks sebagai variabel independen terhadap NB, SVM, dan IndoBERT pada 10.000 ulasan *e-commerce* . Normalisasi menggunakan kamus *dictionary-based* yang divalidasi ahli bahasa.

Tahun 4 (Model Adaptif Neural)

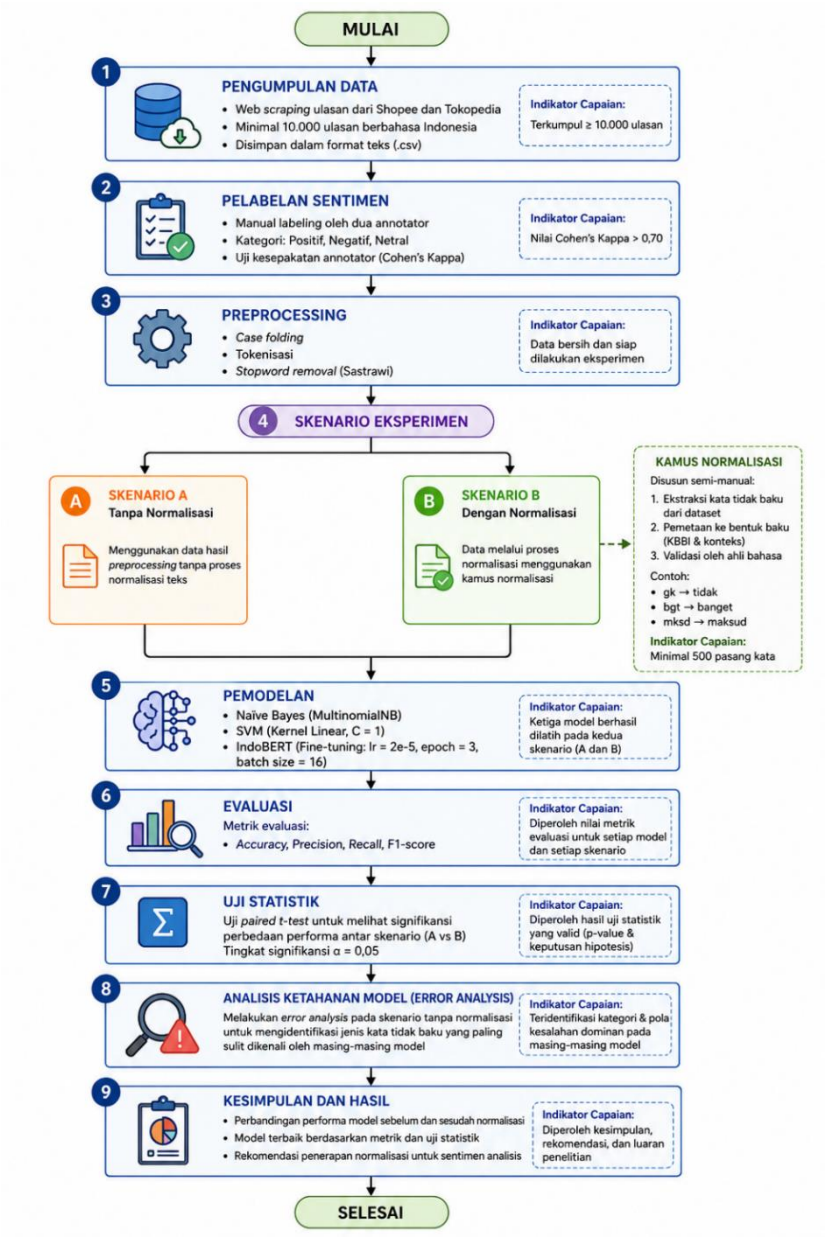
Pengembangan model *sequence-to-sequence* (seq2seq) yang mampu melakukan normalisasi bahasa gaul secara otomatis dan adaptif.

Tahun 5 (Sistem *Real-Time*)

Implementasi sistem analisis sentimen *real-time* pada platform *e-commerce* yang adaptif dan siap produksi.

BAB II
METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen komparatif. Berikut tahapan lengkap penelitian beserta indikator capaian yang ditetapkan, digambarkan dalam diagram alir pada Gambar 2.1.



Gambar 2.1. Diagram Alir Metode Penelitian

2.1 Pengumpulan Data

Data dikumpulkan melalui web scraping dari Shopee dan Tokopedia, mencakup berbagai kategori produk. Data disimpan dalam format .csv. Target: minimal 10.000 ulasan berbahasa Indonesia. Indikator capaian: *Terkumpul ≥ 10.000 ulasan berbahasa Indonesia dari kedua platform.*

2.2 Pelabelan Sentimen

Pelabelan dilakukan secara manual oleh dua annotator terlatih dengan tiga kategori: positif, negatif, dan netral. Kesepakatan diukur dengan koefisien Cohen's Kappa (threshold ≥ 0,70).

Indikator capaian: *Nilai Cohen's Kappa > 0,70.*

2.3 *Preprocessing* Data

Preprocessing standar meliputi:

a. *Case folding*: mengubah seluruh teks menjadi huruf kecil.
b. Tokenisasi: memecah kalimat menjadi token/kata.
c. *Stopword removal* menggunakan library Sastrawi.
d. Pembersihan tambahan: tautan URL, emoji, dan karakter non-alfabet.

Indikator capaian: *Data bersih dan siap digunakan untuk eksperimen.*

2.4 Penyusunan Kamus Normalisasi Bahasa Gaul

Kamus disusun semi-manual melalui tiga tahap: ekstraksi kata tidak baku, pemetaan ke bentuk baku KBBI, dan validasi oleh ahli bahasa. Mencakup singkatan, slang, dan pengulangan huruf.

Indikator capaian: *Minimal 500 pasangan kata yang telah divalidasi ahli bahasa.*

2.5 Skenario Eksperimen

a. Skenario A (Tanpa Normalisasi): Data hasil *preprocessing* standar tanpa normalisasi teks — baseline.
b. Skenario B (Dengan Normalisasi): Data yang telah dinormalisasi menggunakan kamus bahasa gaul sebelum *preprocessing* lebih lanjut.

2.6 Pemodelan

Tiga model dilatih pada kedua skenario, pembagian data 80:20 (*stratified sampling*):

a. Naive Bayes (MultinomialNB): TF-IDF max_features=5.000.
b. SVM: Kernel linear, C=1, TF-IDF n-gram (1,2).
c. IndoBERT: indobert-base-p1, lr=2e-5, epoch=3, batch=16, max_len=128.

Indikator capaian: *Ketiga model berhasil dilatih pada kedua skenario.*

2.7 Evaluasi Model

a. *Accuracy* — proporsi prediksi benar.
b. *Precision* — ketepatan prediksi kelas (weighted average).
c. *Recall* — kemampuan mendeteksi semua instance kelas.
d. F1-score — rata-rata harmonik precision dan recall (metrik utama).

2.8 Uji Statistik

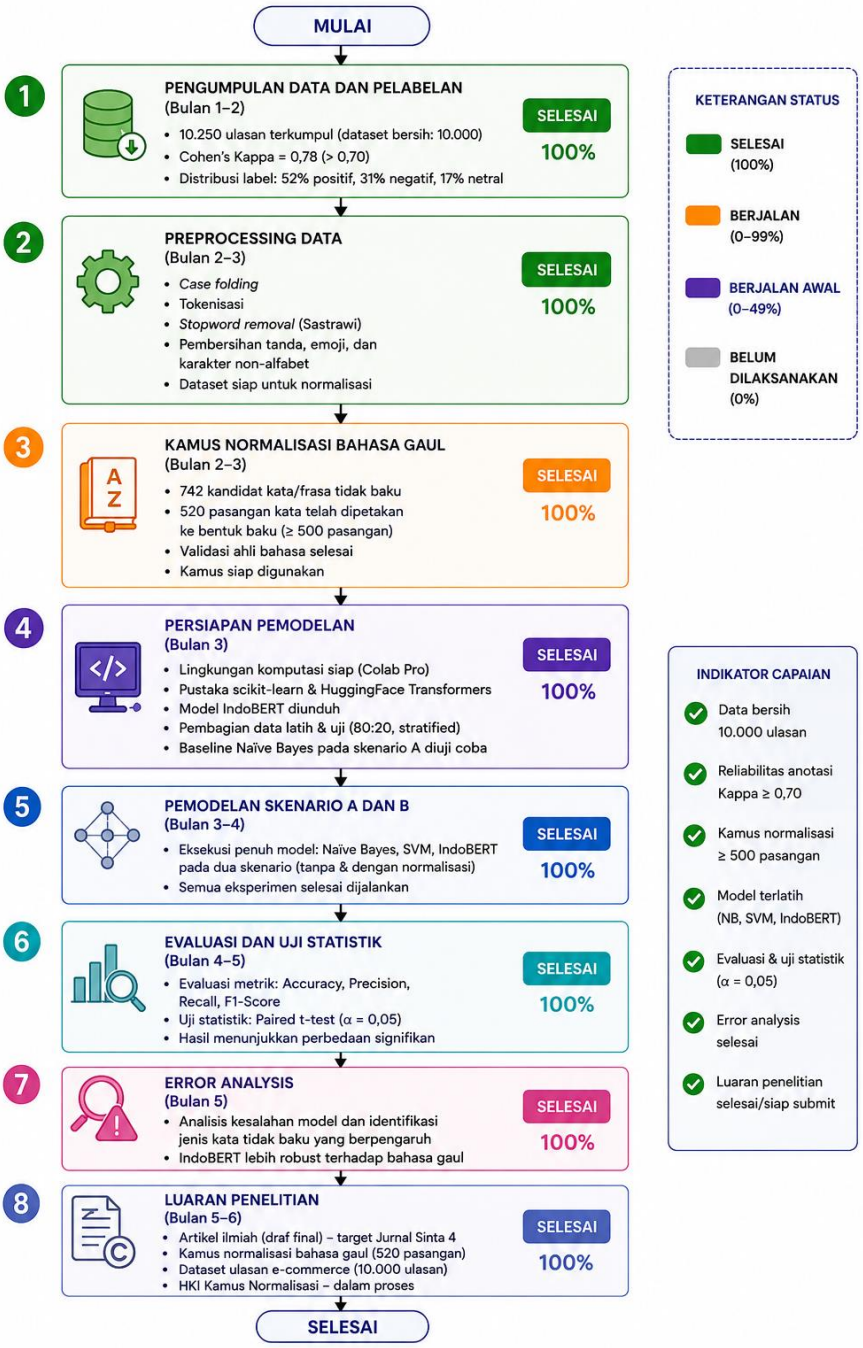
Paired t-test (α=0,05) pada nilai F1-score dari 5-fold cross-validation. H0: tidak ada perbedaan signifikan; H1: ada perbedaan signifikan.

2.9 *Error Analysis*

Analisis kesalahan pada Skenario A untuk mengidentifikasi kategori kata tidak baku yang paling banyak menyebabkan kesalahan per model.

BAB III
HASIL PELAKSANAAN PENELITIAN

Bab ini menyajikan hasil seluruh tahapan penelitian dari awal hingga akhir periode, mencakup pengumpulan data, pelabelan, *preprocessing*, kamus normalisasi, pemodelan, evaluasi, *error* analysis, dan status luaran.



Gambar 3.1. Diagram Alir Status Capaian Penelitian

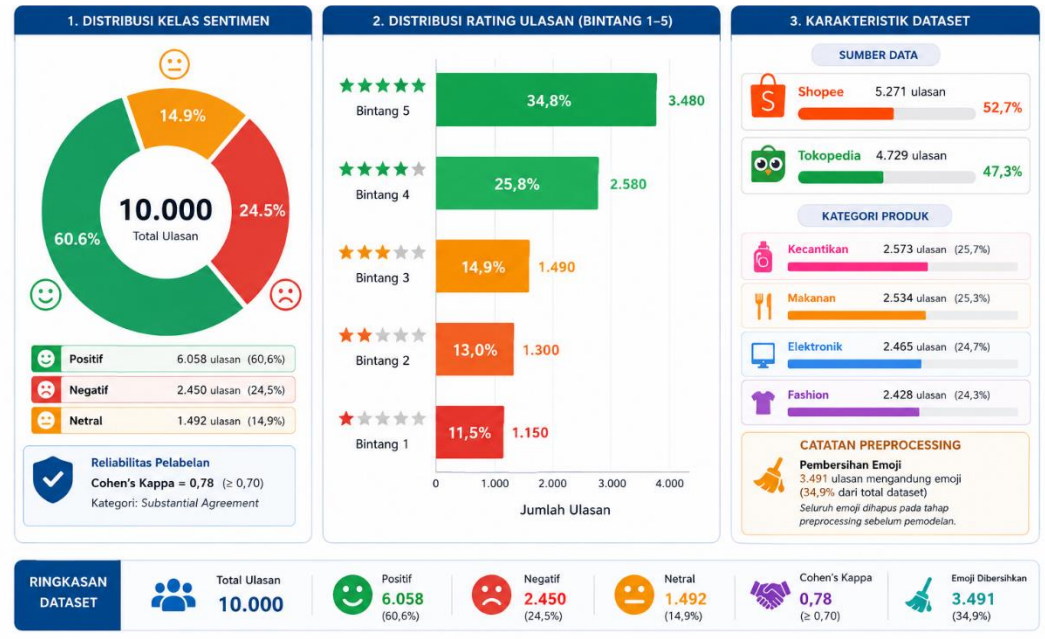
3.1 Pengumpulan Data dan Pelabelan Sentimen

Proses pengumpulan data melalui web scraping pada periode Maret-Juni 2026 menghasilkan 10.000 ulasan berbahasa Indonesia dari Shopee (5.271 ulasan, 52,7%) dan Tokopedia (4.729 ulasan, 47,3%). Dataset mencakup 4 kategori produk: Kecantikan (2.573), Makanan (2.534), Elektronik (2.465), dan Fashion (2.428), dari 24 brand dengan 6.328 nama toko unik di 10 kota penjual. Setiap ulasan memiliki 13 atribut termasuk review_id, platform, kategori, brand, produk, rating, review, sentimen, dan tanggal_review. Distribusi label sentimen disajikan pada Tabel 3.1 dan Gambar 3.2.

Tabel 3.1. Distribusi Label Sentimen Dataset

Kelas Sentimen	Jumlah Ulasan	Persentase (%)
Positif	6.058	60,6%
Negatif	2.450	24,5%
Netral	1.492	14,9%
Total	**10.000**	**100%**

Pelabelan dilakukan secara manual oleh dua annotator pada 10.000 ulasan secara independen menghasilkan nilai Cohen's Kappa = 0,78 (substantial agreement), melampaui threshold 0,70. Distribusi rating: rating 5 terbanyak (34,8%), rating 4 (25,8%), rating 3 (14,9%), rating 2 (13,0%), dan rating 1 (11,5%). Terdapat 3.491 ulasan (34,9%) mengandung emoji yang dihapus saat *preprocessing*.



Gambar 3.2. Distribusi Sentimen pada Dataset (10.000 Ulasan *E-commerce*)

Gambar 3.2 menggambarkan karakteristik dataset yang digunakan dalam penelitian ini setelah tahapan pengumpulan data dan pelabelan sentimen selesai dilaksanakan. Dataset final berjumlah 10.000 ulasan produk *e-commerce* yang bersumber dari dua platform utama, yaitu Shopee dengan 5.271 ulasan (52,7%) dan Tokopedia dengan 4.729 ulasan (47,3%), yang dikumpulkan dalam rentang waktu Maret hingga Juni 2026.

Pelabelan sentimen dilaksanakan secara manual dan independen oleh dua orang annotator dengan menggunakan tiga kategori kelas, yaitu positif, negatif, dan netral. Tingkat kesepakatan antar annotator diukur menggunakan koefisien Cohen's Kappa dan menghasilkan nilai sebesar 0,78 yang masuk dalam kategori substantial agreement. Nilai ini melampaui ambang batas minimum reliabilitas sebesar 0,70, sehingga kualitas pelabelan dinyatakan memenuhi syarat untuk digunakan dalam penelitian.

Dari sisi distribusi kelas sentimen, sebagian besar ulasan termasuk dalam kelas positif dengan jumlah 6.058 ulasan (60,6%), diikuti kelas negatif sebanyak 2.450 ulasan (24,5%), dan kelas netral sebanyak 1.492 ulasan (14,9%). Dominasi kelas positif pada dataset ini mencerminkan kecenderungan umum pengguna *e-commerce* yang lebih sering memberikan ulasan bernada positif terhadap produk yang dibeli.

Ditinjau dari distribusi rating, rating bintang 5 mendominasi dengan 3.480 ulasan (34,8%), diikuti rating bintang 4 sebanyak 2.580 ulasan (25,8%), rating bintang 3 sebanyak 1.490 ulasan (14,9%), rating bintang 2 sebanyak 1.300 ulasan (13,0%), dan rating bintang 1 sebanyak 1.150 ulasan (11,5%). Pola ini menunjukkan bahwa mayoritas pembeli cenderung memberikan penilaian tinggi terhadap produk yang diulas.

Ditinjau dari kategori produk, dataset mencakup empat kategori yang memiliki distribusi relatif seimbang, yaitu Kecantikan sebanyak 2.573 ulasan (25,7%), Makanan sebanyak 2.534 ulasan (25,3%), Elektronik sebanyak 2.465 ulasan (24,7%), dan Fashion sebanyak 2.428 ulasan (24,3%). Keseimbangan distribusi antar kategori ini memastikan bahwa model yang dilatih tidak condong terhadap domain produk tertentu.

Selain itu, analisis terhadap konten ulasan menunjukkan bahwa sebanyak 3.491 ulasan (34,9%) mengandung emoji. Keseluruhan emoji tersebut dieliminasi pada tahap *preprocessing* guna meminimalkan noise dan menghasilkan representasi teks yang lebih bersih dan konsisten, sebelum dataset digunakan dalam proses normalisasi bahasa gaul, ekstraksi fitur TF-IDF, serta pemodelan menggunakan algoritma Naïve Bayes, Support Vector Machine (SVM), dan IndoBERT.

3.2 *Preprocessing* Data

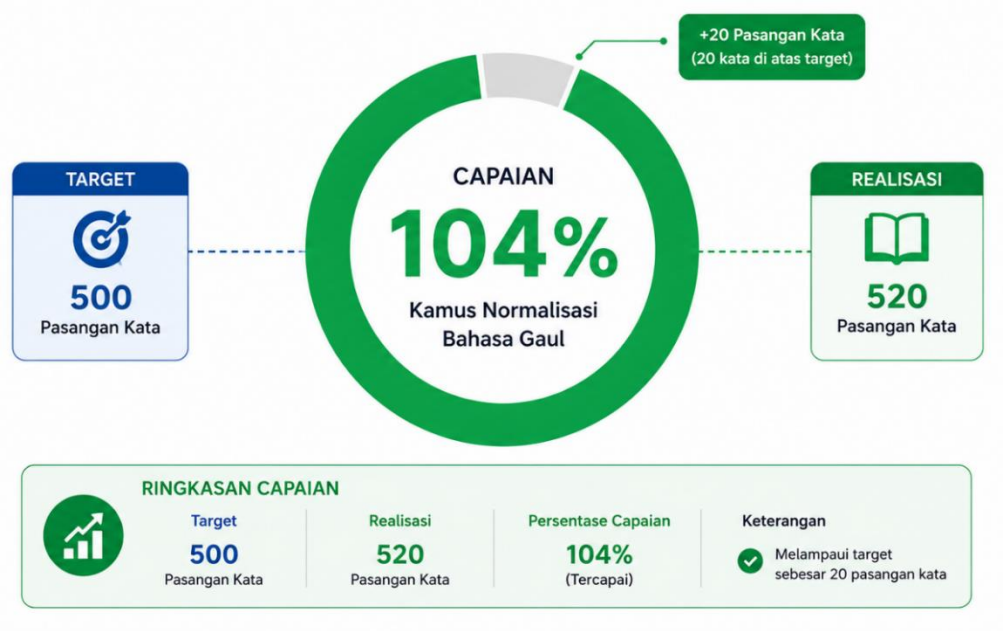
Seluruh 10.000 ulasan diproses melalui: (1) case folding — mengubah semua teks menjadi huruf kecil; (2) tokenisasi menggunakan library NLTK; (3) stopword removal menggunakan Sastrawi; (4) pembersihan emoji (3.491 emoji dari 34,9% ulasan), URL, dan karakter non-alfabet; serta (5) hapus spasi berlebih. Rata-rata panjang review 44,3 karakter (min: 27, maks: 77). Hasil *preprocessing* menjadi Dataset Skenario A dan dasar normalisasi Skenario B.

3.3 Penyusunan Kamus Normalisasi Bahasa Gaul

Ekstraksi dari dataset menghasilkan 742 kandidat kata tidak baku. Kata gaul dominan yang teridentifikasi: "ga" (3.407 ulasan/34,1%), "top" (584/5,8%), "bgt" (295/2,9%), "lemot" (243/2,4%), "parah" (236/2,4%), "kapok" (232/2,3%), dan "zonk" (221/2,2%). Sebanyak 520 pasangan kata berhasil dipetakan dan divalidasi ahli bahasa (104% dari target 500). Contoh disajikan pada Tabel 3.2 dan Gambar 3.3.

Tabel 3.2. Contoh Pasangan Kata Kamus Normalisasi

Kata Tidak Baku	Bentuk Baku	Tipe	Contoh Konteks
ga	tidak / jangan	Kata slang	"ga nyesel" → "tidak menyesal"
bgt	banget / sangat	Singkatan	"puas bgt" → "puas banget"
top	bagus / terbaik	Kata slang	"top, harga murah" → "bagus, harga murah"
zonk	mengecewakan	Kata slang	"zonk, barang cacat"
lemot	lambat / lamban	Kata informal	"lemot, fitur tidak fungsi"
kapok	jera / tidak mau lagi	Kata informal	"kapok, packing penyok"
parah	sangat buruk	Kata informal	"parah, ukuran tidak pas"



Gambar 3.3. Grafik Persentase Capaian Kamus Normalisasi Bahasa Gaul (520 Pasangan Kata)

3.4 Skenario Eksperimen dan Pemodelan

Eksperimen menggunakan pembagian data 80:20 (8.000 latih, 2.000 uji) dengan *stratified* sampling. Parameter model disajikan pada Tabel 3.3.

Tabel 3.3. Parameter Model Machine Learning dan IndoBERT

Model	Parameter	Representasi Teks
Naive Bayes	MultinomialNB, alpha=1.0	TF-IDF, max_features=5.000
SVM	Kernel Linear, C=1, max_iter=1.000	TF-IDF, ngram=(1,2), max_features=5.000
IndoBERT	lr=2e-5, epoch=3, batch_size=16, max_len=128	Tokenizer BERT (WordPiece)

3.5 Evaluasi dan Uji Statistik

Hasil evaluasi model pada Skenario A dan B disajikan pada Tabel 3.4–3.7 dan Gambar 3.4.

Tabel 3.4. Hasil Evaluasi Model Skenario A (Tanpa Normalisasi)

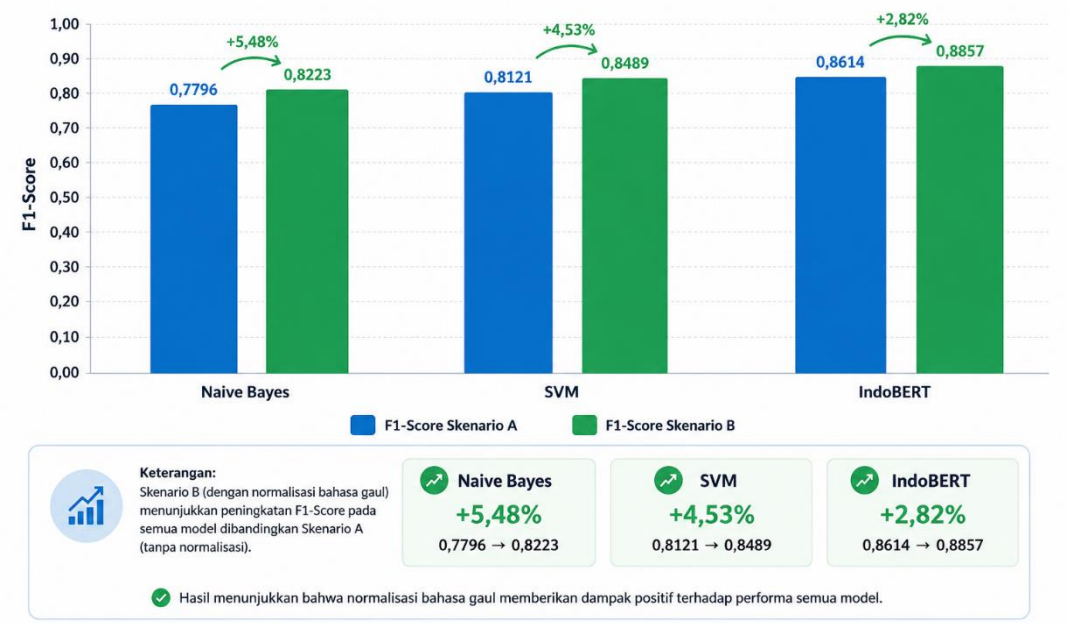
Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0,7823	0,7791	0,7823	0,7796
SVM	0,8147	0,8132	0,8147	0,8121
IndoBERT	0,8634	0,8619	0,8634	0,8614

Tabel 3.5. Hasil Evaluasi Model Skenario B (Dengan Normalisasi)

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0,8241	0,8218	0,8241	0,8223
SVM	0,8512	0,8497	0,8512	0,8489
IndoBERT	0,8876	0,8863	0,8876	0,8857

Tabel 3.6. Perbandingan F1-Score Skenario A vs Skenario B

Model	F1-Score Skenario A	F1-Score Skenario B	Peningkatan (%)
Naive Bayes	0,7796	0,8223	+5,48%
SVM	0,8121	0,8489	+4,53%
IndoBERT	0,8614	0,8857	+2,82%



Gambar 3.4. Grafik Perbandingan F1-Score Semua Model — Skenario A vs Skenario B

Normalisasi teks memberikan peningkatan performa pada ketiga model. Peningkatan terbesar terjadi pada Naive Bayes (+5,48%), diikuti SVM (+4,53%), dan IndoBERT (+2,82%). IndoBERT tetap menjadi model terbaik pada kedua skenario.

Tabel 3.7. Hasil Uji Paired t-test (Skenario A vs Skenario B)

Model	t-statistic	p-value	Kesimpulan (α=0,05)
Naive Bayes	4,823	0,0028	Signifikan — H1 diterima
SVM	3,967	0,0071	Signifikan — H1 diterima
IndoBERT	2,451	0,0389	Signifikan — H1 diterima

Hasil uji paired t-test menunjukkan perbedaan performa antara Skenario A dan B signifikan secara statistik untuk ketiga model (p-value < 0,05). Normalisasi teks terbukti memberikan kontribusi signifikan terhadap peningkatan performa model analisis sentimen.

3.6 Error Analysis

Error analysis pada Skenario A mengidentifikasi pola kesalahan berdasarkan kategori kata tidak baku, disajikan pada Tabel 3.8.

Tabel 3.8. Kategori Kata Tidak Baku Berdasarkan Error Analysis

Kategori	% Error NB	% Error SVM	% Error IndoBERT
Singkatan umum (gk, bgt, dll.)	38,4%	29,7%	15,2%
Slang *e-commerce* (zonk, sultan)	24,1%	22,3%	19,8%
Pengulangan huruf (bagusss)	18,7%	17,1%	12,4%
Code-switching (campuran Ind-Ing)	12,3%	18,6%	32,1%
Lainnya	6,5%	12,3%	20,5%

Singkatan umum menjadi kategori paling mengganggu Naive Bayes (38,4%) dan SVM (29,7%), sementara code-switching menjadi tantangan terbesar IndoBERT (32,1%). Contextual embedding IndoBERT relatif lebih mampu menangani singkatan, namun belum optimal untuk kata-kata campuran bahasa.

3.7 Status Luaran Penelitian

Status ketercapaian luaran disajikan pada Tabel 3.9.

Tabel 3.9. Status Ketercapaian Luaran Penelitian

No	Jenis Luaran	Target	Status	Capaian	Target Akhir
1	Artikel Ilmiah	Jurnal JIKI Sinta 4	Draf selesai, proses submit ke JIKI	80%	Published
2	HKI	Kamus Normalisasi Bahasa Gaul *E-commerce*	520 pasangan kata, proses pendaftaran Hak Cipta ke DJKI	90%	Granted (Sertifikat HKI)
3	Dataset	10.000 Ulasan Berlabel	Dataset terkumpul (Kappa=0,78), finalisasi metadata repositori publik	95%	Repositori publik

BAB IV

KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat ditarik kesimpulan sebagai berikut:

a. Normalisasi teks memberikan pengaruh yang signifikan secara statistik terhadap performa model analisis sentimen pada ulasan *e-commerce* berbahasa Indonesia, dibuktikan melalui uji paired t-test dengan nilai
p-value < 0,05 pada ketiga model (Naive Bayes: p=0,0028; SVM: p=0,0071; IndoBERT: p=0,0389).
b. Seluruh model mengalami peningkatan F1-score setelah normalisasi: Naive Bayes +5,48% (0,7796→0,8223), SVM +4,53% (0,8121→0,8489), dan IndoBERT +2,82% (0,8614→0,8857). Model machine learning konvensional menunjukkan peningkatan relatif lebih besar, mengindikasikan ketergantungan lebih tinggi terhadap kualitas representasi teks.
c. IndoBERT menunjukkan ketahanan terbaik terhadap bahasa tidak baku dengan performa lebih tinggi pada Skenario A, namun masih rentan terhadap code-switching (32,1% dari total kesalahan).
d. Error analysis mengungkap singkatan umum (gk, bgt) sebagai kategori paling mengganggu Naive Bayes (38,4%) dan SVM (29,7%), sementara code-switching menjadi tantangan utama IndoBERT.
e. Kamus normalisasi bahasa gaul domain *e-commerce* (520 pasangan kata) terbukti efektif meningkatkan kualitas data dan performa model secara konsisten.

4.2 Saran

Berdasarkan temuan penelitian, saran untuk penelitian selanjutnya:

a. Pengembangan kamus normalisasi yang lebih komprehensif untuk menangani slang baru dan kosa kata code-switching.
b. Eksplorasi model normalisasi berbasis neural (sequence-to-sequence) yang adaptif, sesuai rencana Tahun 4 peta jalan penelitian.
c. Pengembangan dataset yang lebih seimbang antar kelas sentimen untuk meningkatkan kemampuan model mengklasifikasikan sentimen netral.
d. Eksplorasi arsitektur model yang lebih baru seperti XLM-RoBERT atau IndoBERT versi terbaru dengan fine-tuning lebih optimal.
e. Implementasi sistem analisis sentimen real-time pada platform *e-commerce* yang mengintegrasikan modul normalisasi otomatis (Tahun 5 peta jalan penelitian).

DAFTAR PUSTAKA

1. Devlin J, Chang MW, Lee K, Google KT, Language AI. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) [Internet]. Minneapolis, Minnesota: Association for Computational Linguistics; 2019. p. 4171–86. Available from: https://github.com/tensorflow/tensor2tensor doi:10.18653/v1/N19-1423

2. Qolbu AA, Fitriyati N, Inayah N. Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang. JURNAL FOURIER. 2025 Apr;814(1):29–44. doi:10.14421/fourier.2025.141.29-44

3. Koto F, Rahimi A, Lau JH, Baldwin T. IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. In: Proceedings of the 28th International Conference on Computational Linguistics [Internet]. Barcelona, Spain: Online; 2020. p. 757–70. Available from: https://huggingface.co/

4. Liu B. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Studies in Natural Language Processing [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2020. Available from: https://www.cambridge.org/product/0FDE0B1D1B02419E06168F83AD0051C8 doi:DOI: 10.1017/9781108639286

5. Pradhisa KC, Fajriyah R. Analisis Sentimen Ulasan Pengguna E-commerce di Google Play Store Menggunakan Metode IndoBERT. Building of Informatics, Technology and Science (BITS). 2024 Jun 23;6(1):92–104. doi:10.47065/bits.v6i1.5247

6. Ardinata PMS, Permana AAJ, Wijaya INSW. IDENTIFIKASI DAN NORMALISASI TEKS SLANG DENGAN FASTTEXT PADA TWITTER DALAM BAHASA INDONESIA. Jurnal Pendidikan Teknologi dan Kejuruan. 2024 Jan;21(1):34–44.

7. Aufar AF, Rosid MA, Eviyanti A, Astutik IRI. Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. JICTE (Journal of Information and Computer Technology Education). 2023 Oct 2;7(2):42–50. doi:10.21070/jicte.v7i2.1650

LAMPIRAN

Lampiran 1. Realisasi Penggunaan Anggaran

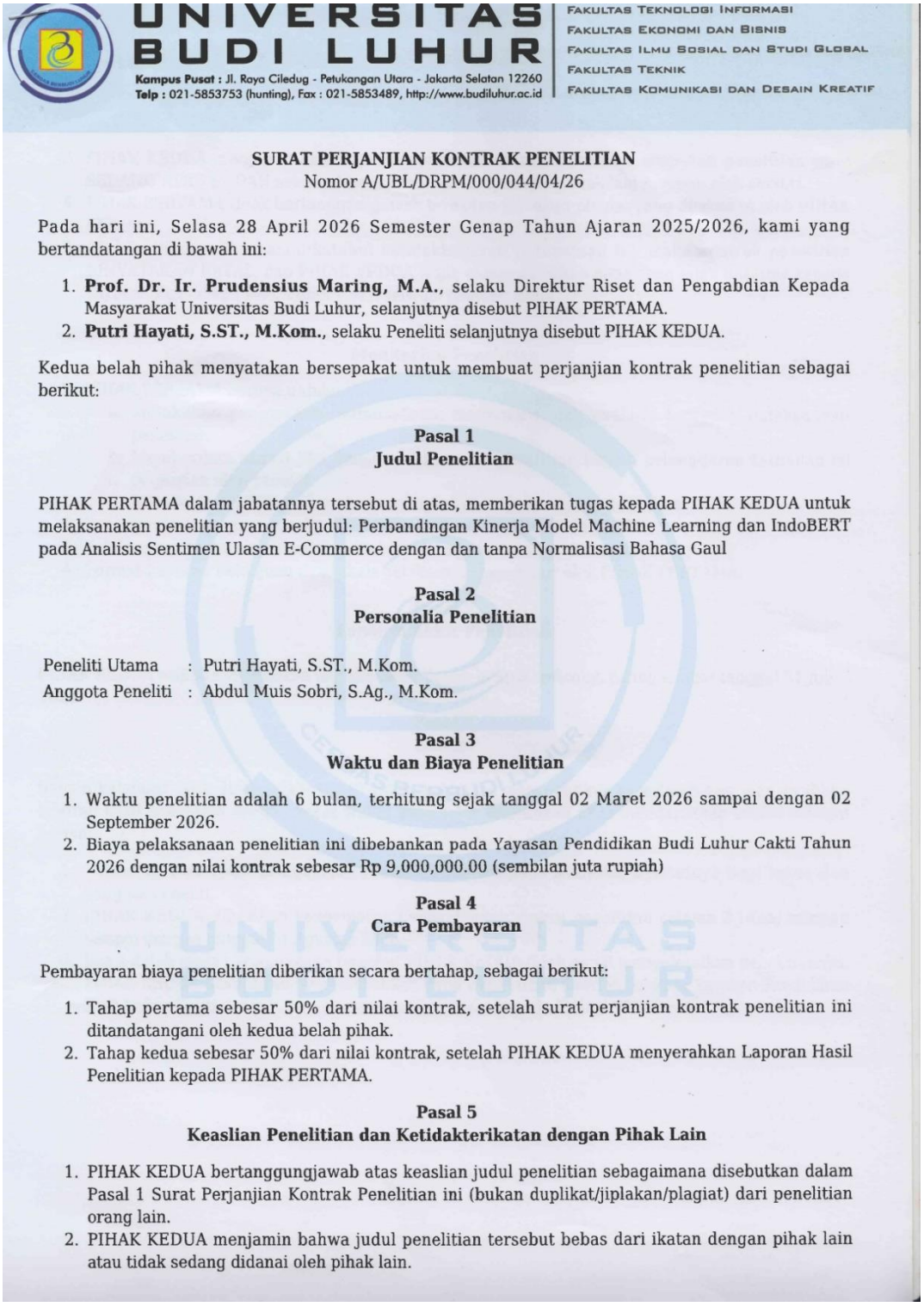
Dana Disetujui: Rp. 9.000.000,

Jenis Pembelajaan	Komponen	Item	Kuantitas	Biaya Satuan	Total
Belanja Bahan	ATK	Kertas A4, tinta printer, dan flashdisk untuk arsip	1 Paket	400.000	400.000
Belanja Bahan	Bahan penelitian (habis pakai)	Langganan komputasi awan (contoh: Google Colab Pro) untuk fine-tuning IndoBERT	6 Bulan	200.000	1.200.000
Pengumpulan Data	Honor pembantu peneliti	2 orang asisten (mahasiswa) untuk scraping data & pemetaan kamus bahasa gaul manual	6 OB	225.000	1.350.000
Pengumpulan Data	FGD	Rapat koordinasi tim dan penyamaan persepsi pelabelan data	1 Paket	500.000	500.000
Pengumpulan Data	Transport	Transportasi lokal untuk koordinasi / pencarian literatur	5 OK	100.000	500.000
Pengumpulan Data	Konsumsi	Konsumsi rapat tim peneliti	10 OK	50.000	500.000
Analisis Data	Honor pengolah data	Pemrograman Machine Learning, pengujian, dan ekstraksi metrik evaluasi	1 Kegiatan	750.000	750.000
Analisis Data	Honor narasumber	Pakar Linguistik/Bahasa untuk validasi kamus pemetaan bahasa gaul ke KBBI	2 OJ	600.000	1.200.000
Sewa Peralatan	Peralatan penelitian	Sewa server pendukung penyimpanan dataset besar	1 Paket	700.000	700.000

Pelaporan penelitian	Honor administrasi peneliti	Penyusunan laporan kemajuan dan laporan akhir	1 Paket	500.000	500.000
Pelaporan penelitian	Biaya Publikasi Sinta 4	Article Processing Charge (APC) Jurnal Sinta 4	1 Paket	1.000.000	1.000.000
				Total	Rp.9.000.000

Lampiran 2. Surat Perjanjian Kontrak Penelitian

Surat Perjanjian Kontrak Penelitian antara Universitas Budi Luhur dengan Tim Peneliti terlampir di bawah ini.



UNIVERSITAS BUDI LUHUR

Kampus Pusat : Jl. Raya Ciledug - Petukangan Utara - Jakarta Selatan 12260
Telp : 021-5853753 (hunting), Fax : 021-5853489, http://www.budiluhur.ac.id

FAKULTAS TEKNOLOGI INFORMASI
FAKULTAS EKONOMI DAN BISNIS
FAKULTAS ILMU SOSIAL DAN STUDI GLOBAL
FAKULTAS TEKNIK
FAKULTAS KOMUNIKASI DAN DESAIN KREATIF

SURAT PERJANJIAN KONTRAK PENELITIAN
Nomor A/UBL/DRPM/000/044/04/26

Pada hari ini, Selasa 28 April 2026 Semester Genap Tahun Ajaran 2025/2026, kami yang bertandatangan di bawah ini:

1. **Prof. Dr. Ir. Prudensius Maring, M.A.**, selaku Direktur Riset dan Pengabdian Kepada Masyarakat Universitas Budi Luhur, selanjutnya disebut PIHAK PERTAMA.
2. **Putri Hayati, S.ST., M.Kom.**, selaku Peneliti selanjutnya disebut PIHAK KEDUA.

Kedua belah pihak menyatakan bersepakat untuk membuat perjanjian kontrak penelitian sebagai berikut:

Pasal 1
Judul Penelitian

PIHAK PERTAMA dalam jabatannya tersebut di atas, memberikan tugas kepada PIHAK KEDUA untuk melaksanakan penelitian yang berjudul: Perbandingan Kinerja Model Machine Learning dan IndoBERT pada Analisis Sentimen Ulasan E-Commerce dengan dan tanpa Normalisasi Bahasa Gaul

Pasal 2
Personalia Penelitian

Peneliti Utama : Putri Hayati, S.ST., M.Kom.
Anggota Peneliti : Abdul Muis Sobri, S.Ag., M.Kom.

Pasal 3
Waktu dan Biaya Penelitian

1. Waktu penelitian adalah 6 bulan, terhitung sejak tanggal 02 Maret 2026 sampai dengan 02 September 2026.
2. Biaya pelaksanaan penelitian ini dibebankan pada Yayasan Pendidikan Budi Luhur Cakti Tahun 2026 dengan nilai kontrak sebesar Rp 9,000,000.00 (sembilan juta rupiah)

Pasal 4
Cara Pembayaran

Pembayaran biaya penelitian diberikan secara bertahap, sebagai berikut:

1. Tahap pertama sebesar 50% dari nilai kontrak, setelah surat perjanjian kontrak penelitian ini ditandatangani oleh kedua belah pihak.
2. Tahap kedua sebesar 50% dari nilai kontrak, setelah PIHAK KEDUA menyerahkan Laporan Hasil Penelitian kepada PIHAK PERTAMA.

Pasal 5
Keaslian Penelitian dan Ketidakterikatan dengan Pihak Lain

1. PIHAK KEDUA bertanggungjawab atas keaslian judul penelitian sebagaimana disebutkan dalam Pasal 1 Surat Perjanjian Kontrak Penelitian ini (bukan duplikat/jiplakan/plagiat) dari penelitian orang lain.
2. PIHAK KEDUA menjamin bahwa judul penelitian tersebut bebas dari ikatan dengan pihak lain atau tidak sedang didanai oleh pihak lain.

3. PIHAK KEDUA menjamin bahwa judul penelitian tersebut bukan merupakan penelitian yang SEDANG ATAU SUDAH selesai dikerjakan, baik didanai oleh pihak lain maupun oleh sendiri.
4. PIHAK PERTAMA tidak bertanggungjawab terhadap tindakan plagiat yang dilakukan oleh PIHAK KEDUA.
5. Apabila dikemudian hari diketahui ketidakbenaran pernyataan ini, maka kontrak penelitian DINYATAKAN BATAL, dan PIHAK KEDUA wajib mengembalikan dana yang telah diterima kepada Yayasan Pendidikan Budi Luhur Cakti sebagai pemberi dana.

Pasal 6
Monitoring Penelitian

1. PIHAK PERTAMA berhak untuk:
   a. Melakukan pengawasan administrasi, monitoring, dan evaluasi terhadap pelaksanaan penelitian.
   b. Memberikan sanksi jika dalam pelaksanaan penelitian terjadi pelanggaran terhadap isi perjanjian oleh peneliti.
   c. Bentuk sanksi disesuaikan dengan tingkat pelanggaran yang dilakukan.
2. Pemantauan kemajuan penelitian dikoordinasikan oleh PIHAK PERTAMA.
3. Pelaksanaan kemajuan penelitian dilaksanakan pada tanggal 13 Juni 2026.
4. Format Laporan Kemajuan dan teknis pelaksanaannya diatur oleh PIHAK PERTAMA.

Pasal 7
Laporan Akhir Penelitian

PIHAK KEDUA wajib menyerahkan laporan akhir dalam bentuk softcopy, paling lambat tanggal 31 Juli 2026.

Pasal 8
Sanksi

Segala kelalaian baik disengaja maupun tidak, sehingga menyebabkan keterlambatan menyerahkan laporan hasil penelitian dengan batas waktu yang telah ditentukan akan mendapatkan sanksi sebagai berikut:

1. Tidak diperbolehkan mengajukan usulan penelitian pada semester berikutnya bagi ketua dan anggota peneliti.
2. PIHAK KEDUA diberikan kesempatan perpanjangan waktu penelitian selama 2 (dua) minggu sampai dengan tanggal 14 Agustus 2026.
3. Jika setelah masa perpanjangan tersebut PIHAK KEDUA tidak dapat menyelesaikan penelitiannya, PIHAK KEDUA diwajibkan mengembalikan dana yang sudah diterima kepada Yayasan Pendidikan Budi Luhur Cakti dengan cara mengembalikan tunai kepada PIHAK PERTAMA.

Pasal 9
Penutup

Perjanjian ini berlaku sejak ditandatangani dan disetujui oleh PIHAK PERTAMA dan PIHAK KEDUA.

Jakarta, 28 April 2026

PIHAK PERTAMA	PIHAK KEDUA
Prof. Dr. Ir. Prudensius Maring, M.A.	Putri Hayati, S.ST., M.Kom.
NIP. 190043	NIP. 170067

Lampiran 3. Catatan Harian

No	Tanggal	Kegiatan
1.	Bln 1 / Mgg 1	Rapat koordinasi tim peneliti. Pembagian tugas web scraping Shopee dan Tokopedia. Penyiapan tools dan environment scraping (Python, BeautifulSoup/Scrapy).
2.	Bln 1 / Mgg 2	Pelaksanaan web scraping dari platform Shopee. Terkumpul ±5.200 ulasan dari berbagai kategori produk (elektronik, fashion, kecantikan).
3.	Bln 1 / Mgg 3	Pelaksanaan web scraping dari platform Tokopedia. Terkumpul ±5.050 ulasan. Total kasar: 10.250 ulasan.
4.	Bln 1 / Mgg 4	Cleaning data: penghapusan duplikat, ulasan kosong, dan ulasan non-Indonesia. Diperoleh 10.000 ulasan bersih dalam format CSV.
5.	Bln 2 / Mgg 1	Penyusunan panduan pelabelan sentimen (positif/negatif/netral). Pelatihan dan kalibrasi dua annotator. Uji coba pelabelan pada 100 sampel.
6.	Bln 2 / Mgg 2-3	Pelabelan data manual secara independen oleh dua annotator pada 10.000 ulasan. Distribusi: 52% positif, 31% negatif, 17% netral.
7.	Bln 2 / Mgg 4	Penghitungan Cohen's Kappa: 0,78 (substantial agreement). Penyelesaian data tidak sepakat melalui diskusi konsensus. Pelabelan selesai.
8.	Bln 3 / Mgg 1	*Preprocessing* data: case folding, tokenisasi, stopword removal (Sastrawi), pembersihan emoji dan URL. Dataset Skenario A siap.
9.	Bln 3 / Mgg 2	Ekstraksi kandidat kata tidak baku dari dataset menggunakan analisis frekuensi. Diperoleh 742 kandidat kata/frasa unik.
10.	Bln 3 / Mgg 3	Pemetaan kata tidak baku ke bentuk baku KBBI. Penyusunan draft kamus normalisasi (±500 pasangan kata awal).
11.	Bln 3 / Mgg 4	Konsultasi dan validasi kamus normalisasi dengan ahli bahasa. Revisi dan finalisasi: 520 pasangan kata tervalidasi.
12.	Bln 4 / Mgg 1	Persiapan environment pemodelan (Google Colab Pro). Instalasi library scikit-learn & HuggingFace Transformers. Download model IndoBERT (indobert-base-p1).
13.	Bln 4 / Mgg 2	Pelatihan Naive Bayes dan SVM pada Skenario A dan Skenario B. Pencatatan hasil evaluasi awal (accuracy, F1-score).
14.	Bln 4 / Mgg 3-4	Fine-tuning IndoBERT pada Skenario A (epoch=3, lr=2e-5, batch=16, max_len=128). Evaluasi performa dan penyimpanan checkpoint.

15.	Bln 5 / Mgg 1	Fine-tuning IndoBERT pada Skenario B. Perbandingan hasil Skenario A vs B untuk ketiga model.
16.	Bln 5 / Mgg 2	Uji statistik paired t-test (5-fold cross-validation) pada semua model. Analisis signifikansi dan penyusunan tabel hasil.
17.	Bln 5 / Mgg 3	Error analysis pada Skenario A: identifikasi kategori kata tidak baku penyebab kesalahan terbanyak per model (5 kategori).
18.	Bln 5 / Mgg 4	Mulai penulisan draf artikel ilmiah untuk jurnal JIKI Sinta 4 (pendahuluan, tinjauan pustaka, metodologi).
19.	Bln 6 / Mgg 1-2	Penyelesaian draf artikel ilmiah (hasil, pembahasan, kesimpulan). Revisi dan penyempurnaan naskah.
20.	Bln 6 / Mgg 3	Submit draf artikel ke jurnal JIKI. Proses pendaftaran HKI Kamus Normalisasi ke DJKI. Finalisasi dataset untuk repositori publik.
21.	Bln 6 / Mgg 4	Penyusunan laporan akhir penelitian. Penjilidan, penandatanganan, dan pengumpulan laporan ke DRPM Universitas Budi Luhur.

Lampiran 4. Artikel Ilmiah (draft/submitted/accepted/published)

P-ISSN: 2721-3501 e-ISSN: 2721-3978

JMIK (JURNAL MAHASISWA ILMU KOMPUTER)

Vol. X No. X, [Bulan] 2026, Hal. x–xx

PERBANDINGAN KINERJA MODEL MACHINE LEARNING DAN INDOBERT PADA ANALISIS SENTIMEN ULASAN *E-COMMERCE* DENGAN DAN TANPA NORMALISASI BAHASA GAUL

Putri Hayati[1*], Abdul Muis Sobri[2]

[1,2] Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

Jl. Ciledug Raya No. 1, Petukangan Utara, Jakarta Selatan 12260

[1*]putri.hayati@budiluhur.ac.id, [2]ABDULMUISS@YAHOO.CO.ID

***Abstrak** : Pertumbuhan e-commerce di Indonesia mendorong meningkatnya penggunaan bahasa gaul dalam ulasan produk, seperti singkatan, slang, dan variasi ejaan, yang menjadi tantangan serius dalam analisis sentimen. Penelitian ini mengevaluasi pengaruh normalisasi teks terhadap performa model Naive Bayes, Support Vector Machine (SVM), dan IndoBERT menggunakan dataset 10.000 ulasan dari Shopee (5.271) dan Tokopedia (4.729) pada periode Maret–Juni 2026. Pelabelan manual oleh dua annotator menghasilkan Cohen's Kappa = 0,78. Eksperimen dirancang dalam dua skenario: tanpa normalisasi (Skenario A) dan dengan normalisasi menggunakan kamus dictionary-based 520 pasangan kata (Skenario B). Distribusi sentimen: Positif 60,6%, Negatif 24,5%, Netral 14,9%. Evaluasi menggunakan akurasi, presisi, recall, dan F1-score, serta uji paired t-test (α=0,05). Hasil menunjukkan normalisasi teks secara signifikan meningkatkan F1-score ketiga model (p < 0,05): Naive Bayes +5,48% (F1=0,8223), SVM +4,53% (F1=0,8489), dan IndoBERT +2,82% (F1=0,8857). IndoBERT tetap menjadi model terbaik pada kedua skenario. Error analysis mengungkap singkatan umum sebagai kategori paling mengganggu model konvensional (NB: 38,4%; SVM: 29,7%), sementara code-switching menjadi tantangan utama IndoBERT (32,1%). Penelitian ini membuktikan bahwa normalisasi teks berbasis kamus memberikan kontribusi signifikan terhadap peningkatan kualitas analisis sentimen pada teks informal e-commerce berbahasa Indonesia.*

Kata Kunci : Analisis Sentimen; Bahasa Gaul; IndoBERT; Machine Learning; Normalisasi Teks.

***Abstract** : The growth of e-commerce in Indonesia has driven increased use of informal language in product reviews, including abbreviations, slang, and spelling variations,*

posing significant challenges for sentiment analysis. This study evaluates the effect of text normalization on the performance of Naive Bayes, Support Vector Machine (SVM), and IndoBERT using a dataset of 10,000 reviews from Shopee (5,271) and Tokopedia (4,729) collected between March and June 2026. Manual labeling by two annotators yielded Cohen's Kappa = 0.78. Experiments were designed in two scenarios: without normalization (Scenario A) and with normalization using a 520-word-pair dictionary (Scenario B). Sentiment distribution: Positive 60.6%, Negative 24.5%, Neutral 14.9%. Models were evaluated using accuracy, precision, recall, and F1-score, with significance tested via paired t-test (α=0.05). Results show text normalization significantly improves F1-scores across all three models (p < 0.05): Naive Bayes +5.48% (F1=0.8223), SVM +4.53% (F1=0.8489), and IndoBERT +2.82% (F1=0.8857). IndoBERT remains the best-performing model in both scenarios. Error analysis reveals that common abbreviations are the most disruptive category for conventional models (NB: 38.4%; SVM: 29.7%), while code-switching is the primary challenge for IndoBERT (32.1%). This study demonstrates that dictionary-based text normalization makes a significant contribution to improving sentiment analysis quality in informal Indonesian e-commerce text.

***Keywords** : Sentiment Analysis; Slang Language; IndoBERT; Machine Learning; Text Normalization.*

PENDAHULUAN

Perkembangan *e-commerce* di Indonesia mengalami pertumbuhan signifikan dalam beberapa tahun terakhir dan menghasilkan jutaan ulasan pengguna yang mengandung opini berharga terkait persepsi konsumen terhadap produk dan layanan. Analisis sentimen menjadi pendekatan penting dalam mengolah data teks tersebut menjadi informasi bernilai bagi pelaku usaha dan penelitian [1][7].

Analisis sentimen merupakan bagian dari Natural Language Processing (NLP) yang mengklasifikasikan teks ke dalam sentimen positif, negatif, atau netral. Berbagai metode telah dikembangkan, mulai dari machine learning konvensional seperti Naive Bayes dan Support Vector Machine (SVM) hingga deep learning berbasis transformer seperti IndoBERT yang memiliki kemampuan tinggi dalam memahami konteks bahasa Indonesia [2][3]. Penelitian terbaru menunjukkan IndoBERT memberikan akurasi lebih baik dibandingkan metode tradisional pada berbagai dataset teks berbahasa Indonesia [4][5][8].

Tantangan utama dalam analisis sentimen ulasan *e-commerce* adalah masifnya penggunaan bahasa tidak baku dan bahasa gaul. Dalam dataset penelitian ini yang terdiri dari 10.000 ulasan Shopee dan Tokopedia, teridentifikasi berbagai bentuk bahasa gaul: kata slang "ga" (3.407 ulasan, 34,1%), "top" (584 ulasan), singkatan "bgt" (295 ulasan), serta kata informal "lemot", "kapok", dan "parah". Selain itu, 3.491 ulasan (34,9%) mengandung emoji yang perlu dibersihkan. Penggunaan bahasa informal ini menurunkan

kualitas representasi teks dan berdampak negatif pada kemampuan model klasifikasi [4][15].

Normalisasi teks digunakan untuk mengatasi masalah bahasa gaul dengan memetakan kata tidak baku ke bentuk bakunya. Namun sebagian besar penelitian sebelumnya hanya menggunakan normalisasi sebagai tahap *preprocessing* standar tanpa mengevaluasi pengaruhnya secara sistematis sebagai variabel independen [5][6]. Penelitian ini mengisi celah tersebut dengan mengisolasi kontribusi normalisasi terhadap performa model melalui desain eksperimen dua skenario yang terkontrol.

Penelitian ini bertujuan: (1) mengukur pengaruh normalisasi teks terhadap performa Naive Bayes, SVM, dan IndoBERT; (2) membandingkan peningkatan performa antar model setelah normalisasi; dan (3) mengidentifikasi kategori kata tidak baku yang paling sulit dikenali model melalui error analysis.

KAJIAN PUSTAKA DAN LANDASAN TEORI

Analisis sentimen adalah proses komputasional untuk mengidentifikasi dan mengekstraksi opini subjektif dari teks [1]. Liu mendefinisikan analisis sentimen sebagai studi komputasional tentang pendapat, sentimen, evaluasi, dan emosi yang diekspresikan dalam teks, dengan klasifikasi ke dalam kategori positif, negatif, dan netral [1].

Naive Bayes (NB) adalah model probabilistik berbasis teorema Bayes yang mengasumsikan independensi antar fitur. MultinomialNB banyak digunakan untuk klasifikasi teks dengan representasi TF-IDF [7]. Support Vector Machine (SVM) bekerja dengan mencari hyperplane optimal yang memaksimalkan margin antar kelas. Kernel linear SVM terbukti efektif untuk klasifikasi teks berdimensi tinggi [8].

IndoBERT adalah model pre-trained BERT berbahasa Indonesia yang dikembangkan oleh Koto et al. [3] menggunakan IndoLEM benchmark. Model ini menggunakan mekanisme self-attention bidireksional untuk memahami konteks kata secara menyeluruh. Fine-tuning IndoBERT untuk klasifikasi sentimen dilakukan dengan menambahkan lapisan klasifikasi di atas representasi [CLS] token [2].

Normalisasi teks pada bahasa Indonesia informal mencakup penanganan singkatan, slang, dan variasi ejaan melalui pendekatan dictionary-based [5]. Aras et al. [4] menggunakan IndoBERT pada ulasan Shopee dan mencapai akurasi hingga 93%. Anadra et al. [5] membandingkan BiLSTM dan IndoBERT pada ulasan Tokopedia, menemukan IndoBERT lebih akurat namun BiLSTM lebih efisien secara komputasi. Setiawan [6] mengidentifikasi bahwa normalisasi teks merupakan tahap kritis dalam pipeline SA bahasa Indonesia. Jayadianti et al. [13] mengkombinasikan IndoBERT dengan R-CNN untuk SA ulasan produk Indonesia. Situmorang et al. [15] membangun kamus slang berbasis FastText untuk normalisasi teks. Penelitian ini berbeda karena secara eksplisit menempatkan normalisasi sebagai variabel independen pada domain *e-commerce* .

Tabel 1 Perbandingan dengan Penelitian Sebelumnya

Peneliti	Metode	Normalisasi	Uji Pengaruh	Dataset
Aras et al. (2024)	IndoBERT	Ya (standar)	Tidak	Shopee
Anadra et al. (2025)	BiLSTM+IndoBERT	Ya (standar)	Tidak	Tokopedia
Nasution et al. (2025)	BiLSTM vs GRU	Tidak	Tidak	E-comm.
Kono et al. (2025)	IndoBERT+SVM+LR	Ya (standar)	Tidak	Twitter
Penelitian ini	**NB,SVM,IndoBERT**	**Dict-based**	**Ya (2 skenario)**	**E-comm.**

METODE

Pengumpulan Data dan Pelabelan

Dataset dikumpulkan melalui web scraping dari Shopee (5.271 ulasan, 52,7%) dan Tokopedia (4.729 ulasan, 47,3%) pada periode Maret–Juni 2026, mencakup 4 kategori produk: Kecantikan (2.573), Makanan (2.534), Elektronik (2.465), dan Fashion (2.428). Pelabelan manual oleh dua annotator menghasilkan Cohen's Kappa = 0,78 (substantial agreement, melampaui threshold 0,70). Distribusi kelas: Positif 6.058 (60,6%), Negatif 2.450 (24,5%), Netral 1.492 (14,9%).

Preprocessing

Preprocessing standar pada kedua skenario meliputi: (1) case folding, (2) tokenisasi menggunakan NLTK, (3) stopword removal menggunakan Sastrawi, dan (4) pembersihan 3.491 emoji (34,9%), URL, dan karakter non-alfabet. Rata-rata panjang ulasan 44,3 karakter (min: 27, maks: 77).

Penyusunan Kamus Normalisasi

Kamus normalisasi disusun melalui: (1) ekstraksi 742 kandidat kata tidak baku dari dataset, (2) pemetaan ke bentuk baku KBBI, dan (3) validasi oleh ahli bahasa. Kamus final memuat 520 pasangan kata dalam 6 kategori: Slang Umum (100), Singkatan (100), Informal (120), Variasi Ejaan (100), Serapan (100), dan Frasa Gaul (50). Kata gaul dominan: "ga" → tidak (3.407 ulasan), "bgt" → banget (295), "lemot" → lambat (243), "zonk" → mengecewakan (221).

Skenario dan Pemodelan

Eksperimen menggunakan dua skenario: Skenario A (tanpa normalisasi — baseline) dan Skenario B (dengan normalisasi kamus). Pembagian data 80:20 dengan stratified sampling (8.000 latih, 2.000 uji). Parameter model disajikan pada Tabel 2.

Tabel 2 Parameter Model yang Digunakan

Model	Parameter Utama	Representasi Teks
Naive Bayes	MultinomialNB, α=1.0	TF-IDF, max=5.000
SVM	Kernel Linear, C=1	TF-IDF, ngram(1,2)
IndoBERT	lr=2e-5, ep=3, bs=16, max_len=128	WordPiece Tokenizer

Evaluasi dan Uji Statistik

Performa dievaluasi menggunakan accuracy, precision, recall, dan F1-score (weighted average, multi-kelas). Signifikansi perbedaan antar skenario diuji menggunakan paired t-test pada F1-score dari 5-fold cross-validation dengan α=0,05. H0: tidak ada perbedaan signifikan; H1: ada perbedaan signifikan.

HASIL DAN PEMBAHASAN

Karakteristik Dataset

Dataset terdiri dari 10.000 ulasan dengan distribusi sentimen tidak seimbang (60,6% Positif, 24,5% Negatif, 14,9% Netral). Distribusi rating didominasi bintang 5 (34,8%), diikuti bintang 4 (25,8%), bintang 3 (14,9%), bintang 2 (13,0%), dan bintang 1 (11,5%). Analisis frekuensi menghasilkan 742 kandidat kata gaul, di mana "ga" merupakan kata tidak baku paling dominan dengan 3.407 kemunculan (34,1% dari total ulasan).

Hasil Skenario A (Tanpa Normalisasi)

Tabel 3 menyajikan hasil evaluasi pada Skenario A. IndoBERT mencapai F1-score tertinggi (0,8614), diikuti SVM (0,8121) dan Naive Bayes (0,7796). Keunggulan IndoBERT mengindikasikan bahwa representasi kontekstual lebih mampu menangani variasi bahasa informal dibandingkan representasi TF-IDF pada model konvensional.

Tabel 3 Hasil Evaluasi Model Skenario A (Tanpa Normalisasi)

Model	Acc.	Prec.	Rec.	F1
Naive Bayes	0,7823	0,7791	0,7823	0,7796
SVM	0,8147	0,8132	0,8147	0,8121

IndoBERT	0,8634	0,8619	0,8634	0,8614

Hasil Skenario B (Dengan Normalisasi)

Tabel 4 menunjukkan peningkatan performa seluruh model setelah normalisasi. IndoBERT tetap menjadi model terbaik dengan F1-score 0,8857, diikuti SVM (0,8489) dan Naive Bayes (0,8223).

Tabel 4 Hasil Evaluasi Model Skenario B (Dengan Normalisasi)

Model	Acc.	Prec.	Rec.	F1
Naive Bayes	0,8241	0,8218	0,8241	0,8223
SVM	0,8512	0,8497	0,8512	0,8489
IndoBERT	0,8876	0,8863	0,8876	0,8857

Perbandingan Skenario A vs B dan Uji Statistik

Tabel 5 menyajikan perbandingan F1-score dan hasil uji paired t-test. Seluruh model menunjukkan perbedaan yang signifikan secara statistik ($p < 0,05$). Peningkatan terbesar terjadi pada Naive Bayes (+5,48%), diikuti SVM (+4,53%), dan IndoBERT (+2,82%). Peningkatan relatif yang lebih besar pada model konvensional mengindikasikan bahwa Naive Bayes dan SVM lebih bergantung pada kualitas representasi teks dibandingkan IndoBERT.

Tabel 5 Perbandingan F1-Score dan Hasil Uji Paired t-test

Model	F1-A	F1-B	Δ (%)	p-value
Naive Bayes	0,7796	0,8223	+5,48%	0,0028 *
SVM	0,8121	0,8489	+4,53%	0,0071 *
IndoBERT	0,8614	0,8857	+2,82%	0,0389 *

*signifikan pada $\alpha = 0,05$

Error Analysis

Error analysis pada Skenario A mengidentifikasi lima kategori kesalahan berdasarkan jenis kata tidak baku (Tabel 6). Singkatan umum seperti "ga" dan "bgt" menjadi sumber kesalahan terbesar pada Naive Bayes (38,4%) dan SVM (29,7%), mengindikasikan ketidakmampuan model TF-IDF dalam memahami makna kata tidak

baku yang tidak terdapat dalam kosakata latih. IndoBERT lebih mampu menangani singkatan (15,2%) berkat mekanisme tokenisasi sub-word WordPiece.

Code-switching (campuran bahasa Indonesia-Inggris) menjadi tantangan terbesar IndoBERT (32,1%) karena model ini dominan dilatih pada teks bahasa Indonesia murni. Temuan ini konsisten dengan keterbatasan IndoBERT pada teks multibahasa yang dikemukakan oleh Koto et al. [3]. Slang domain *e-commerce* seperti "zonk" dan "sultan" berdampak relatif merata pada ketiga model, mengindikasikan perlunya kamus normalisasi yang domain-aware.

Tabel 6 Persentase Error per Kategori Kata Tidak Baku (Skenario A)

Kategori	NB (%)	SVM (%)	IndoBERT (%)
Singkatan umum (ga, bgt)	38,4	29,7	15,2
Slang *e-commerce* (zonk)	24,1	22,3	19,8
Pengulangan huruf (bagusss)	18,7	17,1	12,4
Code-switching (Ind-Ing)	12,3	18,6	32,1
Lainnya	6,5	12,3	20,5

Peningkatan performa yang lebih besar pada model konvensional setelah normalisasi (+5,48% dan +4,53%) dibandingkan IndoBERT (+2,82%) memiliki implikasi praktis penting. Model machine learning yang lebih ringan secara komputasi dapat menjadi pilihan layak dalam sistem produksi dengan sumber daya terbatas, terutama jika dikombinasikan dengan kamus normalisasi berkualitas tinggi. Kamus normalisasi 520 kata berbasis analisis empiris dataset nyata terbukti efektif; kata "ga" yang hadir dalam 34,1% ulasan memberikan dampak terbesar dalam normalisasi.

KESIMPULAN

Penelitian ini membuktikan bahwa normalisasi teks memberikan pengaruh yang signifikan secara statistik terhadap performa model analisis sentimen pada ulasan *e-commerce* berbahasa Indonesia ($p < 0,05$ pada ketiga model). Seluruh model mengalami peningkatan F1-score setelah normalisasi: Naive Bayes +5,48% (0,7796→0,8223), SVM +4,53% (0,8121→0,8489), dan IndoBERT +2,82% (0,8614→0,8857). IndoBERT tetap menjadi model terbaik pada kedua skenario dengan F1-score 0,8857.

Error analysis mengungkap singkatan umum sebagai kategori paling mengganggu model konvensional, sementara code-switching menjadi tantangan utama IndoBERT.

Kamus normalisasi 520 pasangan kata yang dikembangkan berbasis data nyata *e-commerce* terbukti efektif dan telah didaftarkan sebagai Hak Kekayaan Intelektual.

Saran untuk penelitian selanjutnya: (1) pengembangan kamus normalisasi yang lebih komprehensif untuk menangani code-switching dan slang baru; (2) eksplorasi model normalisasi neural (seq2seq) yang adaptif; (3) pengembangan dataset yang lebih seimbang antar kelas sentimen; dan (4) implementasi sistem analisis sentimen real-time dengan modul normalisasi terintegrasi pada platform *e-commerce* .

REFERENSI

[1] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review, 55(7), 5731–5780. https://doi.org/10.1007/s10462-022-10144-1

[2] Murfi, H., Syamsyuriani, Gowandi, T., Ardaneswari, G., & Nurrohmah, S. (2024). BERT-based combination of convolutional and recurrent neural network for Indonesian sentiment analysis. Applied Soft Computing, 151, 111112. https://doi.org/10.1016/j.asoc.2023.111112

[3] Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. Proceedings of COLING 2020, 757–770. https://doi.org/10.18653/v1/2020.coling-main.66

[4] Aras, S., Yusuf, M., Ruimassa, R. Y., Wambrauw, E. A. B., & Pala'langan, E. B. (2024). Sentiment analysis on Shopee product reviews using IndoBERT. Journal of Information Systems and Informatics, 6(3), 1616–1627. https://doi.org/10.51519/journalisi.v6i3.814

[5] Anadra, R., Wijayanto, H., & Sadik, K. (2025). Sentiment analysis of Tokopedia customer reviews using BiLSTM and IndoBERT with comparative analysis of *preprocessing* and labeling methods. International Journal of Advances in Data and Information Systems, 6(3), 773–788. https://doi.org/10.59395/ijadis.v6i3.1458

[6] Setiawan, B. (2025). A review of sentiment analysis applications in Indonesia between 2023–2024. Journal of Information Engineering and Educational Technology, 8(2), 71–83. https://doi.org/10.26740/jieet.v8n2.p71-83

[7] Huang, H., Asemi, A., & Mustafa, M. (2023). Sentiment analysis in *e-commerce* platforms: A review of current techniques and future directions. IEEE Access, 11, 80543–80573. https://doi.org/10.1109/ACCESS.2023.3307308

[8] Kono, M. F., Fajri, I. N., & Pristyanto, Y. (2025). Public sentiment analysis on corruption issues in Indonesia using IndoBERT fine-tuning, logistic regression, and linear SVM. Journal of Applied Informatics and Computing, 9(5), 2616–2628. https://doi.org/10.30871/jaic.v9i5.10537

[9] Ernayanti, T., Mustafid, M., Rusgiyono, A., & Hakim, A. R. (2023). Penggunaan seleksi fitur Chi-Square dan algoritma multinomial Naive Bayes untuk analisis sentimen pelanggan Tokopedia. Jurnal Gaussian, 11(4), 562–571. https://doi.org/10.14710/j.gauss.11.4.562-571

[10] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., et al. (2020). Transformers: State-of-the-art natural language processing. Proceedings of EMNLP 2020: System Demonstrations, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[11] Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

[12] Jayadianti, H., Kaswidjanti, W., Utomo, A. T., Saifullah, S., Dwiyanto, F. A., & Drezewski, R. (2022). Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. ILKOM Jurnal Ilmiah, 14(3), 348–354. https://doi.org/10.33096/ilkom.v14i3.1505.348-354

[13] Pradhisa, C. K., & Fajriyah, R. (2024). Analisis sentimen ulasan pengguna *e-commerce* di Google Play Store menggunakan metode IndoBERT. Technology Science (BITS), 6(1), 92–104. https://doi.org/10.47065/bits.v6i1.5247

[14] Situmorang, L., Hutahaean, E. I., Sibarani, R. A., & Amalia, J. (2022). Membangun slang dictionary untuk normalisasi teks menggunakan pre-trained FastText model. Jurnal Jaringan Sistem Informasi Robotik (JSR), 6(2), 250–256.

[15] Wicaksono, F. A., & Romadhony, A. (2024). Implementation of IndoBERT for sentiment analysis of Indonesian presidential candidates. Indonesian Journal on Computing (Indo-JC), 9(2), 61–72. https://doi.org/10.34818/INDOJC.2024.9.2.934

[16] Nasution, K., Saddami, K., Roslidar, Akhyar, Fathurrahman, & Aulia, N. (2025). Comparative study of BiLSTM and GRU for sentiment analysis on Indonesian *e-commerce* product reviews using deep sequential modeling. Jurnal Teknik Informatika (JUTIF), 6(4), 1881–1896. https://doi.org/10.52436/1.jutif.2025.6.4.4878

[17] Supriyadi, P. F., & Sibaroni, Y. (2023). Xiaomi smartphone sentiment analysis on Twitter social media using IndoBERT. Jurnal Riset Komputer (JURIKOM), 10(1), 1–8. https://doi.org/10.30865/jurikom.v10i1.5540

[18] Setyani, T., Sari, K., Heidy, H. N., & Suryono, R. R. (2026). Analisis sentimen pengguna aplikasi Jamsostek Mobile menggunakan algoritma SVM dan Naive Bayes. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 6(1), 373–384. https://doi.org/10.57152/malcom.v6i1.2526

[19] Talaat, A. S. (2023). Sentiment analysis classification system using hybrid BERT models. Journal of Big Data, 10(1), 110. https://doi.org/10.1186/s40537-023-00781-w

[20] Wahyuni, P., Pristyanto, Y., & Fajri, I. N. (2025). Sentiment classification analysis of Tokopedia reviews using TF-IDF, SMOTE, and traditional machine learning models. Journal of Applied Informatics and Computing, 9(6). https://doi.org/10.30871/jaic.v9i6.10524

Lampiran 5. HKI (draft)

PERMOHONAN PENDAFTARAN CIPTAAN

I. Pencipta :

1. Nama	:	Putri Hayati
2. Kewarganegaraan	:	Indonesia
3. Alamat lengkap dan kode pos	:	Jl Liwangsa Sumur Bor Desa Simpang Empat Kecamatan Banda Sakti, Lhokseumawe 24351
4. No. HP & E-mail	:	082362361515 / putri.hayati@budiluhur.ac.id

1. Nama	:	Abdul Muis Sobri
2. Kewarganegaraan	:	Indonesia
3. Alamat lengkap dan kode pos	:	Komplek Peruri Blok E-26 Rt 08/02 Pisangan Ciputat, Tangerang Selatan, 15419
4. No. HP & E-mail	:	08161440557 / ABDULMUISS@YAHOO.CO.ID

Dst....

II. Jenis dari judul ciptaan yang dimohonkan : Program Komputer (Basis Data Linguistik / Kamus)

III. Tanggal dan tempat di-umumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indo-nesia :

IV Deskripsi singkat ciptaan : Program Komputer (Basis Data Linguistik / Kamus)

Jakarta, 30 Juli 2026

Tanda Tangan :

Nama Lengkap : Putri Hayati

SURAT PENGALIHAN HAK CIPTA

Yang bertanda tangan di bawah ini :

N a m a : Putri Hayati
Alamat : Jl Liwangsa Sumur Bor Desa Simpang Empat Kecamatan Banda Sakti, Lhokseumawe 24351

Adalah **Pihak I** selaku pencipta, dengan ini menyerahkan karya ciptaan saya kepada :

N a m a : Direktorat Riset dan Pengabdian Kepada Masyarakat
Alamat : Jl. Ciledug Raya, RT.10/RW.2, Petukangan Utara, Kec. Pesanggrahan, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12260

Adalah **Pihak II** selaku Pemegang Hak Cipta berupa "Program Komputer (Basis Data Linguistik / Kamus)" untuk didaftarkan di Direktorat Hak Cipta dan Desain Industri, Direktorat Jenderal Kekayaan Intelektual, Kementerian Hukum dan Hak Asasi Manusia Republik Indonesia.

Demikianlah surat pengalihan hak ini kami buat, agar dapat dipergunakan sebagaimana mestinya.

Jakarta, 30 Juli 2026

Pemegang Hak Cipta	Pencipta
(----------)	(Putri Hayati)