



**Optimalisasi IndoBERT untuk Analisis Sentimen Studi Kasus:
Aplikasi MyPertamina di Play Store**

LAPORAN TUGAS AKHIR

DWIKA ERFA DIANSHAH

41521010074

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**



**Optimalisasi IndoBERT untuk Analisis Sentimen Studi Kasus:
Aplikasi MyPertamina di Play Store**

LAPORAN TUGAS AKHIR

DWIKA ERFA DIANSHAH

41521010074

**UNIVERSITAS
MERCU BUANA**

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Dwika Erfa Dianshah

NIM : 41521010074

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Optimalisasi IndoBERT untuk Analisis Sentimen Studi Kasus:
Aplikasi MyPertamina di Play Store

Menyatakan bahwa Laporan Skripsi ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Skripsi saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 19 Juli 2025

UNIVERSITAS
MERCU BUANA



Dwika Erfa Dianshah

HALAMAN PENGESAHAN

Laporan Skripsi ini diajukan oleh:

Nama : Dwika Erfa Dianshah

NIM : 41521010074

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Optimalisasi IndoBERT untuk Analisis Sentimen Studi Kasus:
Aplikasi MyPertamina di Play Store

Telah berhasil dipertahankan pada sidang di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Disahkan oleh:

Pembimbing : Nur Ani, S.T., MMSI.

NIDN : 0310117801

Ketua Penguji : Ilham Nugraha, S.Kom, M.Sc

NIDN : 307098904

Penguji 1 : Anis Cherid, SE, MTI

NIDN : 0328127203

Penguji 2 : Dr. Andi Nugroho, ST, M.Kom

NIDN : 0305098303



Jakarta, 19 Juli 2025
MERCU BUANA

Mengetahui,

Dekan

Ketua Program Studi



Dr. Bambang Jokonowo, S.Si., MTI

NIDN : 0320037002



Dr. Hadi Santoso, S.Kom., M.Kom

NIDN : 0225067701

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan ridho-Nya sehingga penulis dapat menyelesaikan tugas akhir yang merupakan salah satu persyaratan kelulusan Program Studi Strata Satu (S1) Jurusan Teknik Informatika, Universitas Mercu Buana.

Penulis menyadari bahwa tugas akhir ini masih jauh dari sempurna karena kesempurnaan sejatinya hanya milik Tuhan Yang Maha Esa. Oleh karena itu, saran dan masukan yang membangun senantiasa penulis terima dengan senang hati. Berkat dukungan, motivasi, bantuan, bimbingan, dan doa dari banyak pihak, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Andi Adriansyah, M.Eng., selaku Rektor Universitas Mercu Buana.
2. Bapak Dr. Bambang Jokonowo, S.Si., M.T.I., selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Dr. Hadi Santoso, S.Kom., M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Mercu Buana.
4. Ibu Nur Ani, S.T., M.M.S.I., selaku dosen pembimbing tugas akhir yang telah memberikan pengarahan, motivasi, dan menyediakan waktu, tenaga, serta pikiran sehingga pembuatan tugas akhir ini dapat terjadwal dengan baik.
5. Dua orang paling berjasa dalam hidup saya, Bapak Erfai dan Ibu Tarti, yang selalu melangitkan doa-doa baik dan menjadi alasan utama saya mampu menyelesaikan pendidikan ini. Bapak saya, yang menunjukkan perannya melalui kerja kerasnya tanpa banyak kata, dan Ibu saya yang senantiasa hadir dengan ketulusan dan usaha yang tidak pernah mengenal lelah. Terima kasih atas segala pengorbanan dan tulus kasih yang diberikan. Meski Bapak dan Ibu tidak sempat merasakan pendidikan di bangku perkuliahan, namun tetap mampu memberikan yang terbaik, bekerja tanpa lelah, mendoakan, serta memberi perhatian dan dukungan hingga saya mampu menyelesaikan studi dan meraih gelar sarjana. Tugas akhir dan gelar ini saya persembahkan untuk Bapak dan Ibu. Sarjana ini kupersembahkan untuk kalian.

6. Kakak saya, Nanda Erfita Dhaudhia, yang telah menjadi bagian dari perjalanan ini dengan cara yang sederhana namun berarti. Terima kasih atas perhatian dan bantuan yang diberikan di setiap tahap yang saya jalani.
7. Diri saya sendiri, Dwika Erfa Dianshah. Terima kasih telah bertahan, berjuang, dan menyelesaikan apa yang telah dimulai. Terima kasih karena telah mampu berusaha keras dan berjuang sejauh ini, karena telah percaya pada diri sendiri bahwa saya bisa melalui semua ini, dan tak pernah memutuskan menyerah sesulit apa pun proses penyusunan tugas akhir ini. Terima kasih karena telah menyelesaikannya sebaik dan semaksimal mungkin. Ini merupakan pencapaian yang patut dibanggakan untuk diri sendiri. Semoga pencapaian ini menjadi langkah awal menuju versi terbaik dari diri sendiri.
8. Teman-teman Teknik Informatika angkatan 2021, terkhusus dari teman-teman KCI dan Sekte K, yang tak mampu penulis sebutkan satu per satu. Terima kasih atas kebersamaan, bantuan, semangat, dan kebahagiaan yang saling dibagikan selama masa perkuliahan dan penyusunan tugas akhir ini.
9. Terakhir, teruntuk calon pasangan hidup penulis di masa depan. Walau kita belum bertemu, atau mungkin sedang berada di waktu dan tempat yang berbeda, penulis percaya bahwa waktu akan mempertemukan kita pada saat yang tepat. Tugas akhir ini menjadi bukti nyata bahwa tidak ada perempuan mana pun yang mendampingi perjuangan penulis saat menyelesaikan tugas akhir ini. Jika nanti kita bertemu, aku harap kamu tidak harus merasakan perasaan cemburu perih karena nama lain yang ada di sini. Penulis hanya sedang mempersiapkan diri menjadi seseorang yang lebih baik untukmu. Semoga kelak kita akan cepat bertemu.

Akhir kata, penulis berharap semoga Tuhan Yang Maha Esa membalas kebaikan dan selalu mencurahkan rahmat, hidayah, serta panjang umur kepada kita semua. Aamiin.

Terima kasih.

Jakarta, 19 Juli 2025



Dwika Erfa Dianshah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Dwika Erfa Dianshah

NIM : 41521010074

Program Studi : Teknik Informatika

Judul Laporan Skripsi : Optimalisasi IndoBERT untuk Analisis Sentimen Studi Kasus:
Aplikasi MyPertamina di Play Store

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 19 Juli 2025

Yang menyatakan,



Dwika Erfa Dianshah.

ABSTRAK

Nama : Dwika Erfa Dianshah
NIM : 41521010074
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Optimalisasi IndoBERT untuk Analisis Sentimen
Studi Kasus: Aplikasi MyPertamina di Play Store
Dosen Pembimbing : Nur Ani, ST, MMSI

Penelitian ini bertujuan untuk melakukan klasifikasi sentimen pada ulasan pengguna aplikasi MyPertamina menggunakan model IndoBERT dengan pendekatan *semi-supervised learning*. Sebanyak 2355 ulasan dikumpulkan melalui web scraping dan diproses melalui tahapan *preprocessing*, kemudian 1.000 data dilabel secara manual untuk menentukan jumlah data optimal dalam pelatihan model. Hasil terbaik diperoleh pada 800 data berlabel dengan F1-score 98,10%, dan *fine-tuning* model selama 10 epoch menghasilkan akurasi 98,75%. Model ini digunakan untuk melakukan *labeling* otomatis dengan *confidence* tinggi (99,1%) dan distribusi sentimen yang didominasi label negatif (77,5%). Untuk mengatasi ketidakseimbangan, diterapkan random *oversampling* sehingga model akhir mencapai F1-score 0,98 dan performa prediksi yang seimbang terhadap kedua kelas. Hasil ini membuktikan bahwa IndoBERT sangat efektif dan andal untuk analisis sentimen teks Bahasa Indonesia.

Kata kunci: Analisis Sentimen, IndoBERT, Labeling Otomatis, MyPertamina, Fine-Tuning, Klasifikasi, Text Mining, NLP

ABSTRACT

Nama : Dwika Erfa Dianshah
NIM : 41521010074
Program Studi : Teknik Informatika
Judul Laporan Skripsi : Optimalisasi IndoBERT untuk Analisis Sentimen
Studi Kasus: Aplikasi MyPertamina di Play Store
Dosen Pembimbing : Nur Ani, ST, MMSI

This study aims to classify sentiment in user reviews of the MyPertamina application using the IndoBERT model with a semi-supervised learning approach. A total of 2,355 reviews were collected via web scraping and preprocessed, with 1,000 manually labeled samples used to determine the optimal training size. The best performance was achieved using 800 labeled samples, reaching an F1-score of 98.10%, while fine-tuning for 10 epochs resulted in 98.75% accuracy. The model was then used for automatic labeling, yielding a high confidence rate (99.1%) and a dominant negative sentiment (77.5%). To address the class imbalance, random oversampling was applied, and the re-trained model achieved a balanced F1-score of 0.98 across both sentiment classes. These results demonstrate that IndoBERT is highly effective and reliable for sentiment analysis in Indonesian-language text.

Keywords: Sentiment Analysis, IndoBERT, Automatic Labeling, MyPertamina, Fine-Tuning, Classification, Text Mining, NLP

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRAK	vii
ABSTRACT.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Teori Pendukung.....	31
2.2.1 Web Scraping.....	31
2.2.2 Aplikasi MyPertamina	32
2.2.3 Preprocessing Text.....	32
2.2.4 Analisis Sentimen	32
2.2.5 IndoBERT	33
BAB III METODE PENELITIAN	35
3.1 Jenis Penelitian.....	35
3.2 Tahapan Penelitian.....	36
3.3 Metrik Evaluasi	38

3.3.1 Confusion Matrix	38
BAB IV PEMBAHASAN	41
4.1 Teknik Pengumpulan Data	41
4.2 Distribusi Rating	41
4.3 Pre-processing Data	42
4.3.1 Cleaning	43
4.3.2 Case Folding	44
4.3.3 Normalisasi	46
4.3.4 Tokenizing	47
4.4 Eksperimen Jumlah Data Labeling Manual	48
4.4.1 Tujuan dan Skenario Eksperimen	48
4.4.2 Classification Report.....	49
4.4.3 Confusion Matrix Analysis.....	51
4.4.4 Hasil Performa Model.....	56
4.4.5 Penentuan Data Optimal	58
4.5 Eksperimen Skenario Fine-Tuning Model IndoBERT	59
4.5.1 Konfigurasi Skenario Eksperimen	61
4.5.2 Classification Report.....	63
4.5.3 Analisis Confusion Matrix.....	66
4.5.4 Hasil Performa Skenario Eksperimen.....	71
4.5.5 Visualisasi Perbandingan Performa	72
4.5.6 Penentuan Parameter Optimal.....	76
4.6 Final Training Model	77
4.6.1 Konfigurasi Final Training	80
4.6.2 Proses Training Model Final.....	82
4.6.3 Evaluasi Performa Model Final	83
4.6.4 Analisis Confusion Matrix.....	85
4.6.5 Penyimpanan Model dan Dokumentasi Hasil.....	87
4.7 Labeling Otomatis.....	89
4.7.1 Setup dan Persiapan Model Untuk Labeling	93
4.7.2 Justifikasi Labeling Otomatis	95
4.7.3 Deteksi dan Persiapan Data Input	97

4.7.4 Eksekusi Proses Labeling Otomatis.....	99
4.7.5 Analisis dan Interpretasi Hasil Labeling.....	100
4.7.6 Visualisasi Hasil Labeling	103
4.7.7 Penyimpanan dan Dokumentasi Hasil	107
4.7.8 Evaluasi Kualitas dan Rekomendasi.....	108
4.8 Visualisasi WordCloud	118
4.8.1 WordCloud Seluruh Data.....	119
4.8.2 WordCloud Sentimen Positif	120
4.8.3 WordCloud Sentimen Negatif	122
BAB V KESIMPULAN DAN SARAN.....	124
5.1 Kesimpulan	124
5.2 Saran	124
DAFTAR PUSTAKA.....	126
LAMPIRAN.....	129



DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	6
Tabel 3. 1 Confusion Matrix untuk Klasifikasi Sentimen	39
Tabel 4. 1 Data Hasil Cleaning	43
Tabel 4. 2 Data Hasil Case Folding	45
Tabel 4. 3 Data Hasil Normalisasi Teks	46
Tabel 4. 4 Hasil Eksperimen pada Berbagai Jumlah Data	56
Tabel 4. 5 Hasil Performa Skenario Eksperimen	71
Tabel 4. 6 Distribusi Data Yang Digunakan Dalam Final Training	81
Tabel 4. 7 Progress Training Loss Per-Epoch Tabel	82
Tabel 4. 8 Hasil Evaluasi Model Final	84
Tabel 4. 9 Dokumentasi File Output	88
Tabel 4. 10 Hasil Validasi Konfigurasi Model untuk Labeling Otomatis	95
Tabel 4. 11 Proses deteksi dan validasi data input	98
Tabel 4. 12 Konfirmasi Persiapan Data Loader	98
Tabel 4. 13 Output Progress Labeling dan Waktu Eksekusi	99
Tabel 4. 14 Statistik Confidence Score	100
Tabel 4. 15 Output Distribusi Label Prediksi	102
Tabel 4. 16 Output Kategorisasi Confidence Level	102
Tabel 4. 17 Output Cross-Tabulation Label vs Confidence Level	103
Tabel 4. 18 Daftar File Hasil Labeling dan Dokumentasi	107
Tabel 4. 19 Output Ringkasan Akhir Proses Labeling	109
Tabel 4. 20 Hasil Quality Check dan Rekomendasi	109
Tabel 4. 21 Distribusi Awal Label dan Rasio Imbalance	110
Tabel 4. 22 Ringkasan Oversampling dan Perubahan Distribusi Data	111
Tabel 4. 23 Perbandingan Jumlah Data Sebelum dan Sesudah Balancing	113
Tabel 4. 24 Pembagian Dataset dan Distribusi Label per Split	113
Tabel 4. 25 Setup Model dan Konfigurasi Training	114
Tabel 4. 26 Hasil Training dan Deteksi Overfitting	115
Tabel 4. 27 Classification Report pada Test Set	116
Tabel 4. 28 Top 10 Kata Populer Keseluruhan Data	119
Tabel 4. 29 Top 10 Kata Populer Sentimen Positif	121

Tabel 4. 30 Top 10 Kata Populer Sentimen Negatif	123
--	-----



DAFTAR GAMBAR

Gambar 1. 1 Rating Penilaian Aplikasi MyPertamina di Google Play Store	2
Gambar 3. 1 Diagram Alur Tahapan Penelitian.....	36
Gambar 4. 1 Proses Web Scraping.....	41
Gambar 4. 2 Distribusi Rating Bulan Oktober 2024.....	42
Gambar 4. 3 Hasil Tokenizing	47
Gambar 4. 4 Classification Report Eksperimen 300 Data	49
Gambar 4. 5 Classification Report Eksperimen 500 Data	50
Gambar 4. 6 Classification Report Eksperimen 800 Data	50
Gambar 4. 7 Classification Report Eksperimen 1000 Data	51
Gambar 4. 8 Confusion Matrix Eksperimen 1 - 300 Data	52
Gambar 4. 9 Confusion Matrix Eksperimen 2 - 500 Data	53
Gambar 4. 10 Confusion Matrix Eksperimen 3 - 800 Data	54
Gambar 4. 11 Confusion Matrix Eksperimen 4 - 1000 Data	55
Gambar 4. 12 Classification Report Eksperimen 1: Fine-Tuning Ya, Epoch 5....	64
Gambar 4. 13 Classification Report Eksperimen 2: Fine-Tuning Ya, Epoch 10..	64
Gambar 4. 14 Classification Report Eksperimen 3: Fine-Tuning Tidak, Epoch 5	65
Gambar 4. 15 Classification Report Eksperimen 4: Fine-Tuning Tidak, Epoch 10	65
Gambar 4. 16 Confusion Matrix - Skenario 1: Fine-Tuning, 5 Epoch	67
Gambar 4. 17 Confusion Matrix - Skenario 2: Fine-Tuning, 10 Epoch	68
Gambar 4. 18 Confusion Matrix - Skenario 3: Non Fine-Tuning, 5 Epoch	69
Gambar 4. 19 Confusion Matrix - Skenario 4: Non Fine-Tuning, 10 Epoch	70
Gambar 4. 20 Grafik Akurasi per Skenario Eksperimen	72
Gambar 4. 21 Grafik Precision per Skenario Eksperimen	73
Gambar 4. 22 Grafik Recall per Skenario Eksperimen.....	74
Gambar 4. 23 Grafik F1-Score per Skenario Eksperimen	74
Gambar 4. 24 Grafik Perbandingan Semua Metrics per Skenario Eksperimen....	75
Gambar 4. 25 Classification Report Model Final	84
Gambar 4. 26 Confusion Matrix Model Final	86
Gambar 4. 27 Output Distribusi Label Prediksi - Bar Chart.....	104

Gambar 4. 28 Output Visualisasi Distribusi Confidence Score - Histogram dan Box Plot	105
Gambar 4. 29 Output Visualisasi Distribusi Confidence Level - Bar Chart Dengan Gradient.....	106
Gambar 4. 30 Output Heatmap Cross-Tabulation - Heatmap Absolut dan Persentase	107
Gambar 4. 31 Visualisasi Distribusi Label Awal.....	111
Gambar 4. 32 Visualisasi Perbandingan Distribusi Label Sebelum dan Sesudah Balancing	112
Gambar 4. 33 Visualisasi Distribusi Label per Split.....	114
Gambar 4. 34 Visualisasi Training Loss dan Validation Accuracy	116
Gambar 4. 35 Confusion Matrix pada Test Set.....	117
Gambar 4. 36 Distribusi Prediksi Model: Sebelum (kiri) dan Sesudah (kanan) Balancing	118
Gambar 4. 37 WordCloud Keseluruhan Data	119
Gambar 4. 38 WordCloud Sentimen Positif	121
Gambar 4. 39 WordCloud Sentimen Negatif.....	122

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi.....	129
Lampiran 2 Curriculum Vitae (CV)	130
Lampiran 3 Surat Pernyataan HAKI	131
Lampiran 4 Sertifikat BNSP.....	133
Lampiran 5 Lembar Revisi Dosen Penguji.....	134
Lampiran 6 Hasil Cek Turnitin.....	136
Lampiran 7 Halaman Persetujuan.....	137

