

**ANALISIS SENTIMEN ULASAN APLIKASI PERSIB
MENGUNAKAN ALGORITMA BERT**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NANA SYARIFUDDIN

21.11.4312

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**ANALISIS SENTIMEN ULASAN APLIKASI PERSIB
MENGUNAKAN ALGORITMA BERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NANA SYARIFUDDIN

21.11.4312

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN ULASAN APLIKASI PERSIB
MENGUNAKAN ALGORITMA BERT**

yang disusun dan diajukan oleh

Nana Syarifuddin

21.11.4312

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Januari 2026

Dosen Pembimbing,


Rifda Faticha Alfa Aziza, S.Kom., M.Kom.
NIK. 190302392

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN ULASAN APLIKASI PERSIB
MENGGUNAKAN ALGORITMA BERT

yang disusun dan diajukan oleh

Nana Syarifuddin

21.11.4312

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Mujiyanto, M.Kom.

NIK. 190302492

Norhikmah, S.Kom., M.Kom.

NIK. 190302245

Rifda Faticha Alfa Aziza, S.Kom., M.Kom.

NIK. 190302392

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Nana Syarifuddin

NIM : 21.11.4312

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen Ulasan Aplikasi PERSIB Menggunakan Algoritma BERT

Dosen Pembimbing : Rifda Faticha Alfa Aziza, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan**, **rumusan** dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 23 Januari 2026

Yang Menyatakan,



Nana Syarifuddin

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas berkat, rahmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "ANALISIS SENTIMEN ULASAN APLIKASI PERSIB MENGGUNAKAN ALGORITMA BERT". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Universitas Amikom Yogyakarta. Penulis menyadari bahwa keberhasilan penyusunan skripsi ini tidak lepas dari bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer Informatika Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi S1 Informatika Universitas Amikom Yogyakarta.
4. Ibu Rifda Faticha Alfa Aziza, S.Kom., M.Kom, selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan yang sangat berharga selama penyusunan skripsi ini.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya di bidang Informatika.

Yogyakarta, 23 Januari 2026

Nana Syarifuddin

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12
2.2.1 Aplikasi PERSIB Bandung	12
2.2.2 Analisis Sentimen	12
2.2.3 Pemrosesan Bahasa Alami (<i>Natural Language Processing</i>)	13
2.2.4 Pra-pemrosesan Data	13
2.2.5 <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	15
2.2.6 Penanganan <i>Imbalanced Data</i>	15
2.2.6 <i>Deep Learning</i>	16

2.2.7	<i>Arsitektur Transformer</i>	16
2.2.8	<i>Bidirectional Encoder Representations from Transformers (BERT)</i>	18
2.2.9	IndoBERT	19
2.2.10	Confusion Matrix	20
BAB III METODE PENELITIAN		22
3.1	Objek Penelitian	22
3.2	Alur Penelitian	22
3.2.1	Pengumpulan Data	23
3.2.2	<i>Exploratory Data Analysis (EDA)</i>	23
3.2.2.1	Distribusi Rating Bintang	23
3.2.2.2	Analisis Frekuensi Kata	23
3.2.3	Pra-pemrosesan Data	24
3.2.3.1	<i>Data Cleaning</i>	24
3.2.3.2	<i>Case Folding</i>	24
3.2.3.3	Normalisasi	25
3.2.4	<i>Data Labeling</i>	25
3.2.5	<i>Data Preparation</i>	25
3.2.5.1	<i>Data Splitting</i>	26
3.2.5.2	<i>Tokenizer & Padding</i>	26
3.2.5.3	<i>Data Loader</i>	27
3.2.6	<i>Modeling</i>	27
3.2.6.1	<i>Load Pre-trained Model</i>	27
3.2.6.2	<i>Setting Hyperparameter</i>	28
3.2.6.3	Skenario 1: Training Data Asli (Imbalanced)	29
3.2.6.4	Skenario 2: Training Data Seimbang (<i>Balanced</i>)	29
3.2.7	<i>Model Evaluation</i>	30
3.3	Alat dan Bahan	30
BAB IV HASIL DAN PEMBAHASAN		31
4.1	Pengumpulan Data	31
4.2	<i>Exploratory Data Analysis (EDA)</i>	31

4.2.1	Distribusi Rating Bintang	32
4.2.2	Analisis Frekuensi Kata	32
4.3	Pra-pemrosesan Data	33
4.3.1	<i>Data Cleaning</i>	33
4.3.2	<i>Case Folding</i>	34
4.3.3	Normalisasi	35
4.4	<i>Data Labeling</i>	35
4.5	<i>Data Preparation</i>	36
4.5.1	<i>Tokenizer & Padding</i>	36
4.5.2	<i>Data Splitting</i>	37
4.5.3	<i>Data Loader</i>	38
4.6	<i>Modeling</i>	38
4.6.1	<i>Load Pre-trained Model</i>	38
4.6.2	<i>Setting Hyperparameter</i>	39
4.6.3	Skenario 1 (Tanpa SMOTE)	39
4.6.4	Skenario 2 (Dengan SMOTE)	40
4.7	<i>Model Evaluation</i>	42
4.7.1	Evaluasi Skenario 1 (Tanpa SMOTE)	42
4.7.2	Evaluasi Skenario 2 (Dengan SMOTE)	43
4.7.3	Analisis Perbandingan Hasil	45
BAB V	PENUTUP	47
5.1	Kesimpulan	47
5.2	Saran	47
REFERENSI		48

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 2.2 Confusion Matrix	20
Tabel 3.1 Setting Hyperparameter	28
Tabel 4.1 Pengumpulan Data	31
Tabel 4.2 Hasil Data Cleaning	34
Tabel 4.3 Hasil Case Folding	34
Tabel 4.4 Hasil Normallisasi	35
Tabel 4.5 Hasil Data Labeling	36
Tabel 4.6 Hasil Data Splitting	38
Tabel 4.7 Riwayat Pelatihan Model Skenario 1	39
Tabel 4.8 Riwayat Pelatihan Model Skenario 2	41
Tabel 4.9 Hasil Classification Report Skenario 1	43
Tabel 4.10 Hasil Classification Report Skenario 2	44



DAFTAR GAMBAR

Gambar 2.1	Arsitektur Transformers	17
Gambar 2.2	Arsitektur BERT / Mekanisme Pre-training BERT	18
Gambar 3.1	Diagram Alur Penelitian	22
Gambar 4.1	Grafik Distribusi Rating Bintang Ulasan Aplikasi PERSIB	32
Gambar 4.2	Grafik 15 Kata Teratas dengan Frekuensi Tertinggi	33
Gambar 4.3	Contoh Hasil Tokenizer & Padding	37
Gambar 4.4	Grafik Perbandingan Training Loss dan Validation Loss Skenario 1	40
Gambar 4.5	Grafik Perbandingan Training Loss dan Validation Loss Skenario 2	41
Gambar 4.6	Hasil Confusion Matrix Skenario 1	42
Gambar 4.7	Hasil Confusion Matrix Skenario 2	44
Gambar 4.8	Grafik Perbandingan Performa Model Skenario 1 vs Skenario 2	45



DAFTAR LAMBANG DAN SINGKATAN

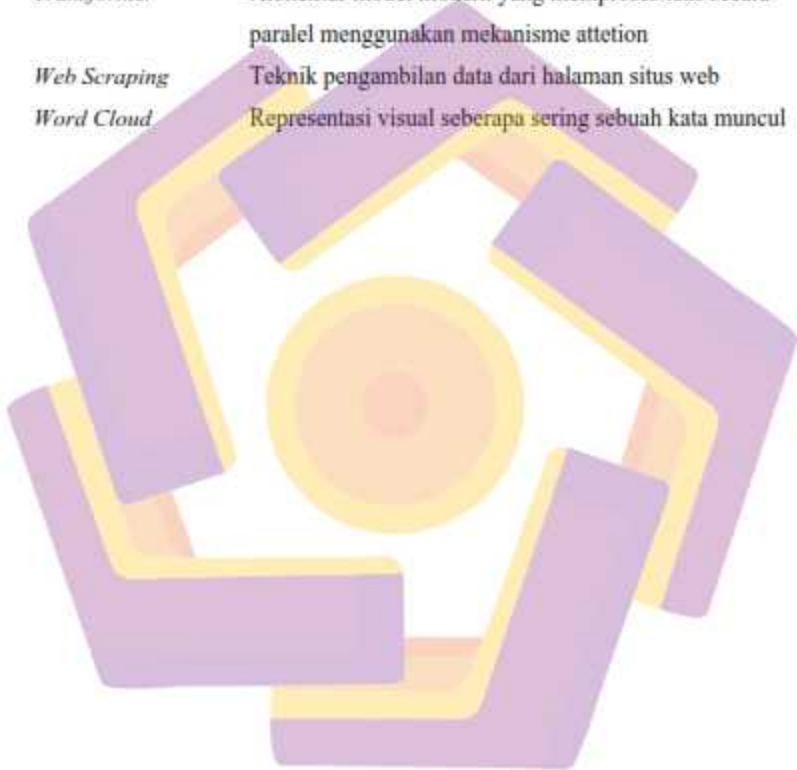


BERT	Bidirectional Encoder Representations from Transformers
CLS	Classification Token
CSV	Comma Separated Values
FN	False Negative
FP	False Positive
MLM	Masked Language Modeling
NLP	Natural Language Processing
NSP	Next Sentence Prediction
PAD	Padding Token
SEP	Separator Token
SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency-Inverse Document Frequency
TN	True Negative
TP	True Positive

DAFTAR ISTILAH

<i>Accuracy</i>	Persentase prediksi benar
<i>Attention</i>	Mekanisme agar model dapat memfokuskan perhatian pada kata yang paling relevan dalam kalimat
<i>Batch Size</i>	Jumlah sampel data yang diproses dalam satu kali iterasi
<i>Benchmark</i>	Standar atau tolok ukur untuk menilai kinerja model
<i>Bobotoh</i>	Pendukung klub sepak bola Persib Bandung
<i>Casefolding</i>	Proses mengubah teks menjadi seragam
<i>Colloquial</i>	Ragam bahasa yang bersifat informal atau sehari-hari
<i>Data Cleaning</i>	Proses membersihkan teks dari emoji, karakter, simbol, dll
<i>Decoder</i>	Komponen model yang bertugas menerjemahkan representasi kode kembali menjadi output
<i>Deep Learning</i>	Metode pembelajaran mesin yang menggunakan jaringan saraf tiruan berlapis banyak
<i>E-commerce</i>	Transaksi jual beli barang yang dilakukan melalui internet
<i>Encoder</i>	Komponen model yang mengubah data input menjadi representasi numerik
<i>Epoch</i>	Satu siklus pelatihan model
<i>F1-Score</i>	Nilai rata-rata harmonik
<i>Fine-tuning</i>	Proses melatih ulang model <i>pre-trained</i> pada data baru
<i>Hyperparameter</i>	Konfigurasi model yang harus diatur secara manual
<i>Labeling</i>	Proses memberikan label pada data
<i>Learning Rate</i>	Parameter pengontrol kecepatan perubahan bobot model
<i>Machine Learning</i>	Sistem komputer yang mampu belajar dari data
<i>Modeling</i>	Proses pelatihan model
<i>Normalisasi</i>	Proses mengubah variasi penulisan kata menjadi seragam
<i>Opinion Mining</i>	Proses komputasi untuk mengidentifikasi opini dari teks
<i>Preprocessing</i>	Tahap persiapan data mentah sebelum diolah oleh model
<i>Precision</i>	Tingkat ketepatan model saat memprediksi kelas positif
<i>Pre-trained Model</i>	Model yang sudah dilatih sebelumnya

<i>Recall</i>	Proporsi data positif yang sebenarnya ada
<i>Slang</i>	Kata atau istilah gaul tidak baku
<i>Softmax</i>	Fungsi matematika untuk mengubah nilai prediksi menjadi persentase probabilitas
<i>Tokenizer</i>	Proses memecah teks panjang menjadi potongan kecil
<i>Transformer</i>	Arsitektur model modern yang memproses teks secara paralel menggunakan mekanisme attention
<i>Web Scraping</i>	Teknik pengambilan data dari halaman situs web
<i>Word Cloud</i>	Representasi visual seberapa sering sebuah kata muncul



INTISARI

Ulasan pengguna pada aplikasi PERSIB di Google Play Store merupakan sumber data krusial yang mencerminkan tingkat kepuasan *bobotoh* terhadap layanan yang diberikan. Penelitian ini bertujuan untuk mengimplementasikan metode *Deep Learning* guna mengklasifikasikan sentimen ulasan tersebut secara akurat ke dalam kategori positif dan negatif, serta mengatasi masalah ketidakseimbangan data (*imbalanced dataset*) yang sering terjadi pada ulasan publik. Pendekatan yang diterapkan adalah *Fine-Tuning* pada model *Pre-trained* IndoBERT varian *indobert-base-p1* yang dikombinasikan dengan teknik *Synthetic Minority Oversampling Technique* (SMOTE) pada data latih. Tahapan penelitian mencakup pengumpulan data, pra-pemrosesan teks, serta pelatihan model dengan skema pembagian data 70:20:10. Berdasarkan hasil pengujian pada skenario terbaik, model menunjukkan performa yang sangat impresif dengan mencatatkan akurasi global sebesar 91% pada data uji. Penerapan SMOTE terbukti signifikan dalam meningkatkan kemampuan deteksi pada kelas minoritas (positif) dengan perolehan *Precision* 0,91 dan *F1-Score* 0,85, sembari mempertahankan kinerja superior pada sentimen negatif dengan *F1-Score* mencapai 0,93. Hasil penelitian ini membuktikan bahwa integrasi IndoBERT dan strategi penyeimbang data mampu mengenali pola bahasa yang kompleks dan menghasilkan klasifikasi yang baik serta objektif.

Kata kunci: Analisis Sentimen, Aplikasi PERSIB, IndoBERT, SMOTE.

ABSTRACT

User reviews of the PERSIB application on the Google Play Store serve as a crucial data source reflecting the satisfaction level of boboto regarding the services provided. This study aims to implement Deep Learning methods to accurately classify these review sentiments into positive and negative categories, while also addressing the issue of imbalanced datasets often encountered in public reviews. The applied approach involves Fine-Tuning the Pre-trained IndoBERT model (indobert-base-p1 variant) combined with the Synthetic Minority Oversampling Technique (SMOTE) on the training data. The research stages included data collection, text pre-processing, and model training utilizing a 70:20:10 data splitting scheme. Based on the testing results from the best scenario, the model demonstrated impressive performance by recording a global accuracy of 91% on the test set. The application of SMOTE proved significant in enhancing detection capabilities for the minority class (positive), achieving a Precision of 0.91 and an F1-Score of 0.85, while maintaining superior performance on negative sentiment with an F1-Score reaching 0.93. These results demonstrate that the integration of IndoBERT and data balancing strategies is capable of recognizing complex language patterns and producing good and objective classifications.

Keyword: Sentiment Analysis, PERSIB Application, IndoBERT, SMOTE