

**PERBANDINGAN ALGORITMA INDOBERT,
INDOBERTWEET, DAN INDOROBERTA UNTUK
KLASIFIKASI EMOSI**

SKRIPSI

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



disusun oleh

EUNIQUE LYDIA STEPHANY

22.11.4545

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**PERBANDINGAN ALGORITMA INDOBERT,
INDOBERTWEET, DAN INDOROBERTA UNTUK
KLASIFIKASI EMOSI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

EUNIQUE LYDIA STEPHANY

22.11.4545

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA INDOBERT, INDOBERTWEET, DAN
INDOROBERTA UNTUK KLASIFIKASI EMOSI**

yang disusun dan diajukan oleh

Eunique Lydia Stephany

22.11.4545

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 26 Februari 2026

Dosen Pembimbing,



Theopilus Bayu Sasongko. S.Kom.,M.Eng
NIK. 190302375

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN ALGORITMA INDOBERT, INDOBERTWEET, DAN INDOROBERTA UNTUK KLASIFIKASI EMOSI

yang disusun dan diajukan oleh

Eunique Lydia Stephany

22.11.4545

Telah dipertahankan di depan Dewan Penguji
pada tanggal 26 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dhani Ariatmanto, S.Kom., M.Kom., Ph.D.
NIK. 190302197



Ferian Fauzi Abdulloh, S.Kom., M.Kom.
NIK. 190302276



Theopilus Bayu Sasongko, S.Kom., M.Eng.
NIK. 190302375



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 26 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Eunique Lydia Stephany
NIM : 22.11.4545

Menyatakan bahwa Skripsi dengan judul berikut:

Perbandingan Algoritma Indobert, Indobertweet, dan Indoroberta Untuk Klasifikasi Emosi

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom.,M.Eng

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, 26 Februari 2026

Yang Menyatakan,



Eunique Lydia Stephany

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada:

1. Kedua orang tua dan saudari tercinta, Dessy Natalia, serta seluruh keluarga besar yang senantiasa memberikan doa, dukungan, serta kepercayaan dalam setiap proses perjalanan akademik penulis.
2. Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng, selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, serta ilmu selama penyusunan skripsi ini.
3. Almamater, yaitu Universitas AMIKOM Yogyakarta yang telah menjadi tempat penulis menimba ilmu.
4. Rekan-rekan seperjuangan, yaitu Bima Pratama, Zahran Nugraha, Zefantio, Andhika Pradana, Oscar Iqbal, serta seluruh kelas 22IF01, yang telah menjadi bagian dari proses pembelajaran dan perjuangan selama masa perkuliahan.
5. Teman-teman terdekat, meskipun berasal dari jurusan yang berbeda, yaitu Prayogi, Ayu, Fignathia, Atmi, Ima, Theresia, dan lainnya yang senantiasa memberikan dukungan dan kebersamaan selama proses studi.
6. Diri penulis sendiri, sebagai bentuk penghargaan atas usaha, ketekunan, dan tanggung jawab dalam menyelesaikan perjalanan studi ini.
7. Untuk Chiyo dan Cheyi, yang dengan kehadiran sederhana telah menemani dan memberi warna selama proses penyusunan skripsi ini.

Semoga karya ini menjadi wujud tanggung jawab akademik serta memberikan manfaat bagi pengembangan ilmu pengetahuan

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “ Perbandingan Algoritma Indobert, Indobertweet, dan Indoroberta untuk Klasifikasi Emosi” dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Universitas Amikom Yogyakarta. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca dan pihak-pihak yang membutuhkan. Dalam proses penyusunan skripsi ini, penulis memperoleh banyak bantuan dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan terima kasih kepada:

1. Prof. Dr. Mohammad Suryanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusrini, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Eli Pujastuti, M.kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan yang sangat berharga selama proses pengerjaan skripsi ini.
5. Tim Dosen Penguji, yang telah memberikan saran, kritik, dan masukan berharga dalam penyempurnaan skripsi ini.
6. Kedua orang tua dan keluarga penulis yang selalu memberikan doa dan dukungan selama proses penyusunan skripsi.
7. Semua pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat disebutkan satu per satu.

Yogyakarta, 17 Februari 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH.....	xiii
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA	8
2.1 Studi Literatur	8

2.2	Dasar Teori.....	14
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian.....	21
3.2	Alur Penelitian	21
3.3	Alat dan Bahan.....	30
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Pengumpulan Data.....	33
4.2	<i>Preprocessing</i> Data.....	33
4.3	<i>Exploratory Data Analysis</i>	36
4.4	<i>Splitting Dataset</i>	41
4.5	<i>Modelling</i>	43
4.6	Tokenisasi	43
4.7	Evaluasi.....	44
4.8	Implementasi.....	54
BAB V PENUTUP		56
5.1	Kesimpulan	56
5.2	Saran	57
REFERENSI		58
LAMPIRAN.....		61

DAFTAR TABEL

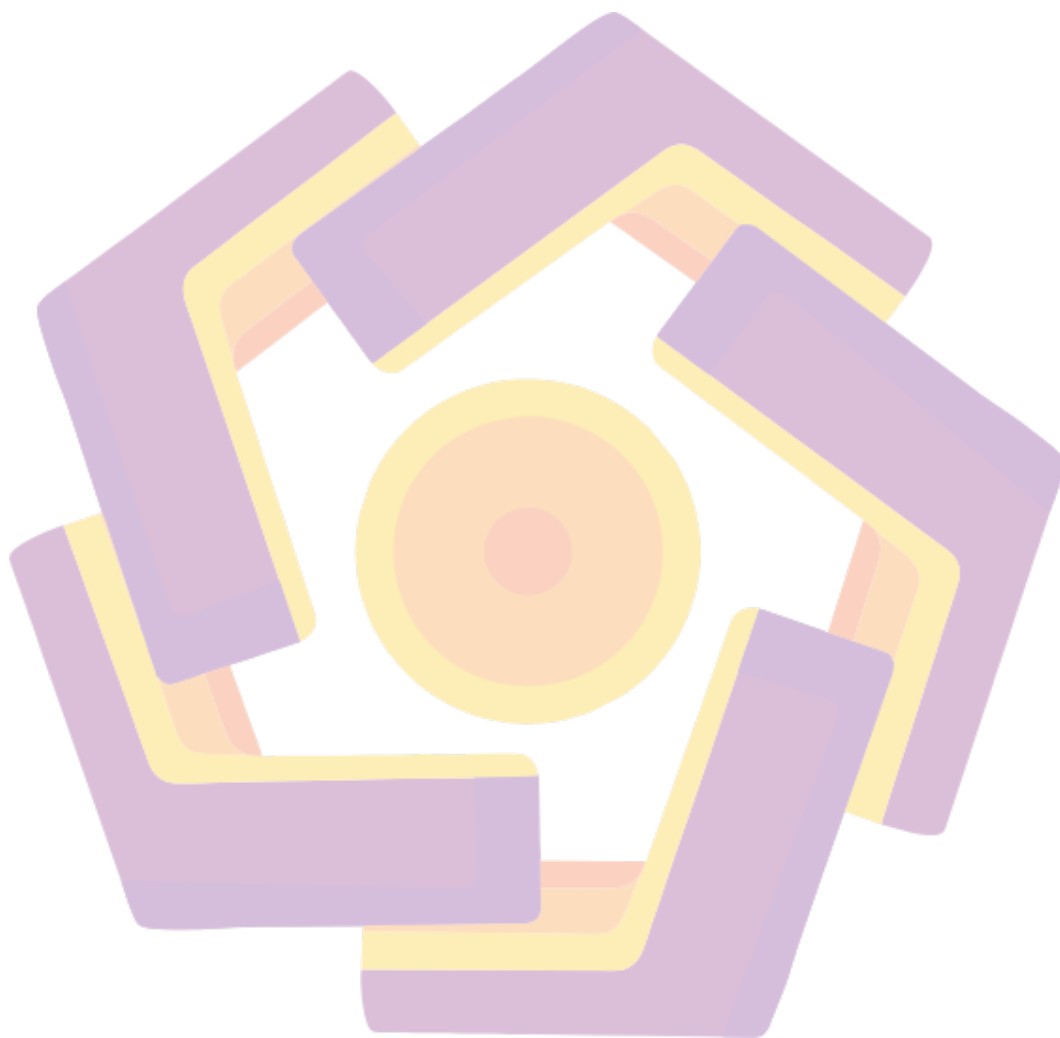
Tabel 2.1 Keaslian Penelitian	11
Tabel 3.1 Konfigurasi Training.....	26
Tabel 3.2 Tabel Perangkat Keras	31
Tabel 3.3 Tabel Perangkat Lunak	32
Tabel 4.1 Index Emosi	45
Tabel 4.2 Evaluasi Metrik Holdout.....	45
Tabel 4.3 Evaluasi Metrik K-fold	46
Tabel 4.4 Waktu Komputasi Holdout	47
Tabel 4.5 Waktu Komputasi K-fold.....	47
Tabel 4.6 Evaluasi Model IndoBERT.....	49
Tabel 4.7 Evaluasi Model IndoBERTweet	50
Tabel 4.8 Evaluasi Model IndoRoBERTa	52

DAFTAR GAMBAR

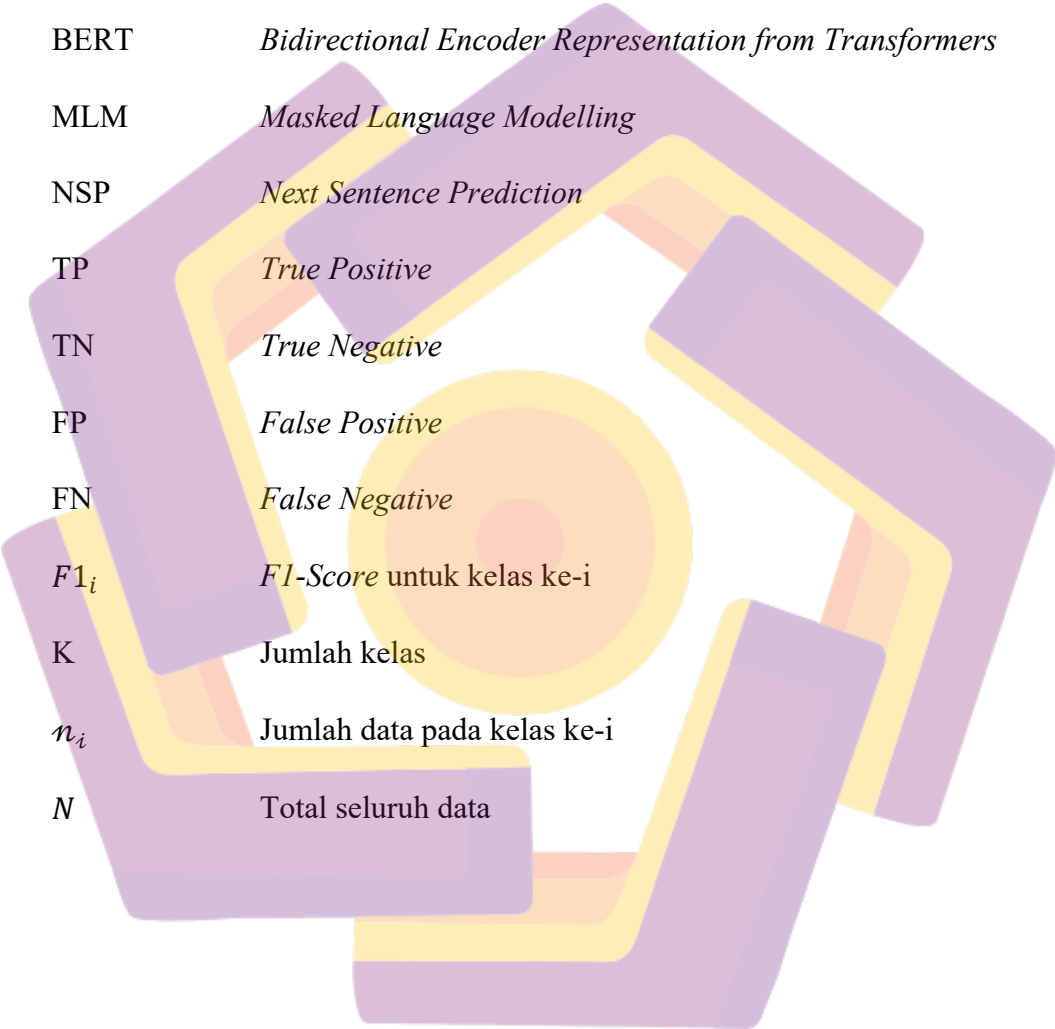
Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Menampilkan <i>dataset</i>	33
Gambar 4.2 Missing Value Checking	34
Gambar 4.3 Overspacing	34
Gambar 4.4 Duplicate Handling	35
Gambar 4.5 Duplikasi setelah ditangani	35
Gambar 4.6 Fungsi <i>clean</i>	36
Gambar 4.7 Cek Baris Kosong	36
Gambar 4.8 Word Cloud Anger	37
Gambar 4.9 Word Cloud Fear	37
Gambar 4.10 Word Cloud Joy	38
Gambar 4.11 Word cloud Love	38
Gambar 4.12 Word cloud Netral	39
Gambar 4.13 Word cloud Sad	39
Gambar 4.14 Grafik Jumlah Data per Label	40
Gambar 4.15 Grafik Rata-Rata Panjang Kalimat	41
Gambar 4.16 Fungsi Splitting Dataset	41
Gambar 4.17 Holdout Splitting	42
Gambar 4.18 K-fold cross validation	42
Gambar 4.19 Loads Models	43
Gambar 4.20 Hyperparameters	43
Gambar 4.21 Hasil Tokenization Sanity Check	44
Gambar 4.22 Tampilan Deployment	54
Gambar 4.23 Contoh Prediksi	55

DAFTAR LAMPIRAN

Lampiran 1 Repository Notebooks	61
---------------------------------------	----



DAFTAR LAMBANG DAN SINGKATAN



NLP	<i>Natural Language Processing</i>
ML	<i>Machine Learning</i>
DL	<i>Deep Learning</i>
BERT	<i>Bidirectional Encoder Representation from Transformers</i>
MLM	<i>Masked Language Modelling</i>
NSP	<i>Next Sentence Prediction</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>
$F1_i$	<i>F1-Score untuk kelas ke-i</i>
K	Jumlah kelas
n_i	Jumlah data pada kelas ke-i
N	Total seluruh data

DAFTAR ISTILAH

Attention Mechanism

mekanisme dalam arsitektur Transformer untuk memperhatikan hubungan antar kata dalam satu kalimat.

Confusion Matrix

tabel evaluasi yang menunjukkan jumlah prediksi benar dan salah dalam proses klasifikasi

Dropout

teknik regularisasi yang menonaktifkan neuron secara acak selama proses pelatihan

Early Stopping

teknik penghentian pelatihan secara otomatis ketika performa validasi tidak lagi meningkat

F1-Score

rata-rata harmonik antara presisi dan recall

Holdout

metode pembagian dataset menjadi data latih, validasi, dan uji dengan proporsi tertentu

K-Fold Cross Validation

metode pembagian dataset menjadi data latih dan data validasi (uji) berdasarkan nilai k.

Macro F1-Score

rata-rata nilai F1 dari seluruh kelas tanpa mempertimbangkan distribusi jumlah data tiap kelas

NLP

bidang kecerdasan buatan yang berfokus pada pengolahan dan pemahaman bahasa manusia

Weight Decay

teknik regularisasi yang membatasi besarnya parameter model selama pelatihan

Dynamic Masking

teknik yang digunakan pada tahap *pretraining* model Transformer, dimana token yang disembunyikan akan berubah ubah setiap kali data diproses dalam *epoch* yang berbeda

Masked Language Modeling (MLM)

teknik *pretraining* BERT dengan menyembunyikan beberapa kata dalam kalimat dan melatih model untuk menebak kata tersebut berdasarkan konteks

Next Sentence Prediction (NSP)

teknik *pretraining* BERT untuk memprediksi apakah dua kalimat saling berhubungan atau merupakan kelanjutan satu sama lain

INTISARI

Perkembangan komunikasi digital yang pesat di Indonesia menyebabkan meningkatnya ekspresi emosi masyarakat dalam bentuk teks pada media sosial. Namun, pengolahan dan pemahaman emosi secara otomatis masih menjadi tantangan karena kompleksitas konteks bahasa dan variasi ekspresi pengguna. Selain itu, penelitian sebelumnya umumnya menggunakan satu model Transformer atau melakukan optimasi model tertentu tanpa melakukan perbandingan langsung antar model bahasa Indonesia secara komprehensif. Permasalahan ini berdampak pada keterbatasan dalam memetakan respons emosional publik yang dapat dimanfaatkan untuk analisis opini, pengambilan keputusan, maupun pengembangan sistem cerdas. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja tiga model Transformer berbasis Bahasa Indonesia, yaitu IndoBERT, IndoBERTweet, dan IndoRoBERTa, dalam tugas klasifikasi enam emosi dasar. Metode yang digunakan meliputi *preprocessing data*, pembagian *dataset* menggunakan skema *holdout* dan *k-fold cross validation*, hingga *modelling algorithm* dengan konfigurasi standar serta *weight decay*, dan *early stopping*. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, dan *macro F1-score*. Hasil penelitian menunjukkan bahwa ketiga model mampu melakukan klasifikasi emosi dengan performa yang baik, terutama IndoBERTweet dengan skema *k-fold* dengan nilai $k = 5$ dan *fold* 2 mendapatkan akurasi sebesar 80,3%, presisi sebesar 80,47%, *recall* sebesar 81,89%, dan *macro f1-score* sebesar 80,99%. Penelitian ini berkontribusi dalam memberikan perbandingan empiris model bahasa Indonesia untuk klasifikasi emosi teks dan dapat dimanfaatkan oleh peneliti, praktisi NLP, maupun pengembang sistem analisis media sosial.

Kata kunci: Klasifikasi Emosi, *Natural Language Processing*, Transformer, Model Berbasis BERT, Bahasa Indonesia

ABSTRACT

The rapid growth of digital communication in Indonesia has led to an increasing expression of emotions through textual content on social media platforms. However, automatically processing and understanding emotions in text remains challenging due to contextual language complexity and variations in user expression. In addition, previous studies generally focused on using a single Transformer model or optimizing a specific model without conducting a comprehensive direct comparison among Indonesian language models. This issue limits the ability to accurately map public emotional responses, which are valuable for opinion analysis, decision-making, and intelligent system development. This study aims to analyze and compare the performance of three Indonesian Transformer-based models, namely IndoBERT, IndoBERTweet, and IndoRoBERTa, in the task of six basic emotion classification. The research methodology includes data preprocessing, dataset splitting using holdout and k-fold cross-validation schemes, and modelling algorithm with standard configurations employing weight decay, and early stopping. Model performance is evaluated using accuracy, precision, recall, and macro F1-score. The results indicate that all three models achieve strong performance in emotion classification, with IndoBERTweet using the k-fold scheme ($k = 5$, fold 2) achieving the best performance, obtaining an accuracy of 80,3%, precision of 80,47%, recall of 81,99%, and macro f1-score of 80,99%. This study contributes an empirical comparison of Indonesian language models for text-based emotion classification and can be utilized by researchers, NLP practitioners, and social media analytics developers.

Keyword: *Emotion Classification, Natural Language Processing, Transformer, BERT-based Models, Indonesian Language.*