

**PERBANDINGAN KINERJA MODEL TRANSFORMER
MONOLINGUAL DALAM ANALISIS SENTIMEN
PUBLIK TERHADAP WEBSITE CORETAX
DI PLATFORM X**

SKRIPSI

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



disusun oleh

ANGGA AIRUL ROHMAN

21.11.4384

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN KINERJA MODEL TRANSFORMER
MONOLINGUAL DALAM ANALISIS SENTIMEN
PUBLIK TERHADAP WEBSITE CORETAX
DI PLATFORM X**

SKRIPSI

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



disusun oleh
ANGGA AIRUL ROHMAN
21.11.4384

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN KINERJA MODEL TRANSFORMER
MONOLINGUAL DALAM ANALISIS SENTIMEN
PUBLIK TERHADAP WEBSITE CORETAX
DI PLATFORM X**

yang disusun dan diajukan oleh

ANGGA AIRUL ROHMAN

21.11.4384

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 19 Agustus 2025

Dosen Pembimbing,



Ike Verawati, S. Kom., M. Kom.
NIK. 190302237

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN KINERJA MODEL TRANSFORMER
MONOLINGUAL DALAM ANALISIS SENTIMEN
PUBLIK TERHADAP WEBSITE CORETAX
DI PLATFORM X

yang disusun dan diajukan oleh

ANGGA AIRUL ROHMAN
21.11.4384

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Agustus 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Dina Maulina, S. Kom., M. Kom.
NIK. 190302250

Mulia Sulistiyono, S. Kom., M. Kom.
NIK. 190302248

Ike Verawati, S. Kom., M. Kom.
NIK. 190302237



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **ANGGA AIRUL ROHMAN**
NIM : **21.11.4384**

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN KINERJA MODEL TRANSFORMER MONOLINGUAL DALAM ANALISIS SENTIMEN PUBLIK TERHADAP WEBSITE CORETAX DI PLATFORM X

Dosen Pembimbing : **Ike Verawati, 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, 19 Agustus 2025

Yang Menyatakan,



ANGGA AIRUL ROHMAN

HALAMAN PERSEMBAHAN

Dengan rasa syukur dan hormat, karya ilmiah ini saya persembahkan kepada:

1. Kedua orang tua saya, Bapak Sutarno dan Ibu Eni Martini atas kasih sayang, doa, dan dukungan moral serta material yang tidak pernah henti dalam setiap langkah hidup saya.
2. Dosen pembimbing saya, Ibu Ike Verawati, M. Kom, yang telah memberikan ilmu, bimbingan, dan arahan selama pengerjaan skripsi saya.
3. Teman-teman seperjuangan, yang telah menjadi sumber semangat dan kebersamaan dalam melewati setiap tantangan akademik.
4. Almamater tercinta, tempat saya tumbuh dan belajar, serta mengembangkan potensi diri di bidang ilmu pengetahuan dan teknologi.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga skripsi yang berjudul “Perbandingan Kinerja Model Transformer Monolingual dalam Analisis Sentimen Publik terhadap Website Coretax di Platform X” ini dapat diselesaikan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis juga menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Ike Verawati, M.Kom. selaku dosen pembimbing yang telah dengan sabar memberikan bimbingan, arahan, serta motivasi selama proses penyusunan skripsi ini.
2. Tim Dosen Penguji Skripsi, yang telah memberikan masukan dan saran konstruktif demi kesempurnaan karya ini.
3. Seluruh dosen dan staf pengajar di Program Studi S1 Informatika, yang telah membekali penulis dengan ilmu pengetahuan selama masa studi.
4. Kedua orang tua, atas doa, dukungan, dan kasih sayang yang tiada henti dalam setiap langkah perjalanan akademik penulis.
5. Teman-teman seperjuangan, atas semangat, dukungan, dan kebersamaan selama masa perkuliahan dan penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang.

Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi kontribusi dalam pengembangan ilmu pengetahuan di bidang teknologi informasi.

Yogyakarta, 16 Agustus 2025



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.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3

BAB II TINJAUAN PUSTAKA	5
2.1 Studi Literatur	5
2.2 Dasar Teori.....	12
2.2.1 Text Mining	12
2.2.2 Analisis Sentimen	12
2.2.3 Media Sosial X.....	12
2.2.4 Natural Language Processing (NLP)	13
2.2.5 Arsitektur Transformer	14
2.2.6 BERT (<i>Bidirectional Encoder Representations from Transformers</i>) 14	
2.2.7 IndoBERT	15
2.2.8 IndoBERTweet	15
2.2.9 IndoRoBERTa.....	16
2.2.10 Python	16
2.2.11 Google Colab	16
2.2.12 Crawling Data	17
2.2.13 Tweet Harvest	17
2.2.14 Preprocessing Data.....	18
2.2.15 Augmentasi Data	18
2.2.16 Confusion Matrix	19
2.2.17 Metrik Evaluasi	20
BAB III METODE PENELITIAN	22
3.1 Alur Penelitian	22
3.3 Alat dan Bahan.....	25

3.3.1	Alat.....	25
3.3.2	Bahan	28
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Pengumpulan Data	30
4.2	Labelling Dataset	32
4.3	Preprocessing	33
4.4	Splitting Dataset.....	38
4.5	Augmentasi Sinonim.....	40
4.6	Implementasi Model	42
4.7	Evaluasi Kinerja Model	43
4.8	Perbandingan hasil	47
BAB V PENUTUP		49
5.1	Kesimpulan	49
5.2	Saran	49
REFERENSI		51
LAMPIRAN.....		58

DAFTAR TABEL

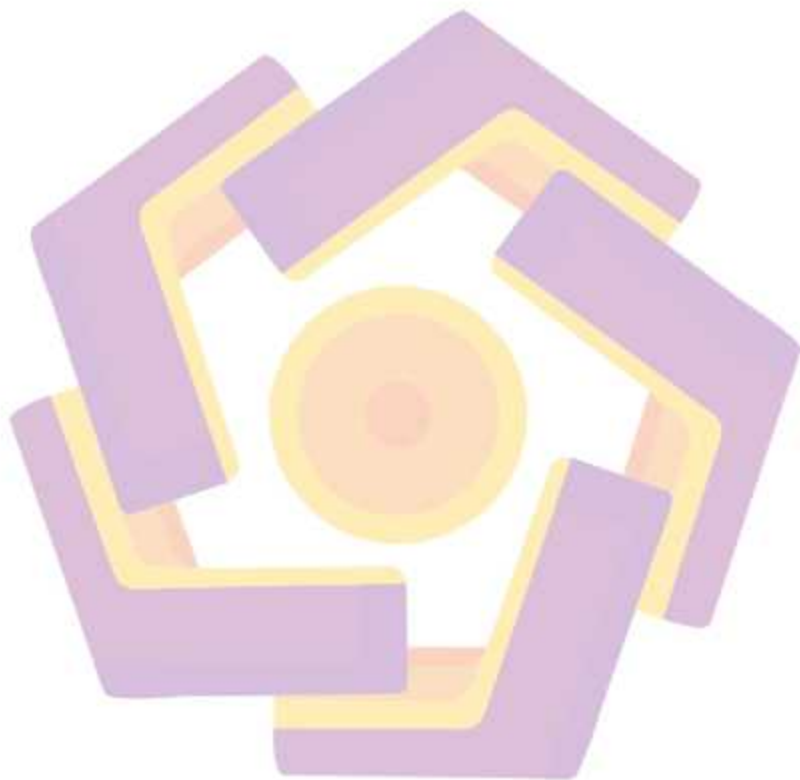
Tabel 3.1 Sampel Dataset	29
Tabel 4.1 Atribut dari Dataset Hasil Crawling	30
Tabel 4.2 Hasil Labelling Dataset.....	33
Tabel 4.3 Hasil Cleaning Data	34
Tabel 4.4 Hasil Case Folding.....	35
Tabel 4.5 Contoh Kata Tidak Baku dari Kamus Alay.....	36
Tabel 4.6 Hasil Normalisasi Kata	37
Tabel 4.7 Hasil Penghapusan Duplikasi Data.....	38
Tabel 4.8 Distribusi Jumlah Label Sentimen (Sebelum Augmentasi).....	40
Tabel 4.9 Hasil Classification Report Macro Average	47

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Hasil Crawling Data	31
Gambar 4.2 Distribusi Jumlah Data per Label Sentimen	32
Gambar 4.4 Distribusi Jumlah Data pada Subset Pelatihan, Validasi, dan Pengujian (Sebelum Augmentasi)	39
Gambar 4.3 Perbandingan Distribusi Data Training Set Sebelum dan Sesudah Augmentasi	41
Gambar 4.5 Konfigurasi Hyperparameter untuk Fine Tuning Model	42
Gambar 4.6 Hasil Confusion Matrix IndoBERT	44
Gambar 4.7 Hasil Confusion Matrix IndoBERTweet	45
Gambar 4.8 Hasil Confusion Matrix IndoRoBERTa	46

DAFTAR LAMPIRAN

Lampiran 1 link <i>google drive</i> program.....	58
--	----



DAFTAR LAMBANG DAN SINGKATAN



API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CSV	Comma-Separated Values
FN	False Negative
FP	False Positive
MLM	Masked Language Modeling
NLTK	Natural Language Toolkit
NLP	Natural Language Processing
NSP	Next Sentence Prediction
RAM	Random Access Memory
TN	True Negative
TAPI	True Positive
URL	Uniform Resource Locator

DAFTAR ISTILAH



Augmentasi Data	Penambahan jumlah data dengan memodifikasi data asli.
Case Folding	Mengubah semua huruf dalam teks menjadi huruf kecil.
Cleaning	Membersihkan teks dari karakter atau elemen tidak relevan
Dataset	Kumpulan data yang digunakan untuk melatih dan menguji model.
Fine-tuning	Penyesuaian ulang pre-trained model dengan data khusus.
Hyperparameter	Parameter yang ditentukan sebelum pelatihan model.
Machine Learning	Algoritma yang memungkinkan komputer belajar dari data
Overfitting	Kondisi model terlalu menyesuaikan data latih sehingga buruk pada data baru
Preprocessing	Tahap awal menyiapkan teks sebelum dianalisis
Tokenizing	Memecah teks menjadi unit kecil seperti kata atau sub-kata
Transformer	Arsitektur model berbasis attention untuk pemrosesan bahasa
Underfitting	Kondisi model gagal mempelajari pola data

INTISARI

Penelitian ini bertujuan untuk membandingkan kinerja tiga model Transformer monolingual berbahasa Indonesia, yaitu *IndoBERT*, *IndoBERTweet*, dan *IndoRoBERTa*, dalam analisis sentimen terhadap opini publik mengenai website Coretax pada platform X. Tahapan penelitian meliputi pengumpulan data, praproses teks, augmentasi data, serta *fine-tuning* terhadap masing-masing model. Evaluasi dilakukan menggunakan *classification report* berdasarkan metrik *precision*, *recall*, *f1-score*, dan akurasi. Hasil penelitian menunjukkan bahwa *IndoBERT* memberikan performa terbaik dibandingkan kedua model lainnya dengan nilai *macro average* tertinggi pada seluruh metrik evaluasi. Temuan ini diharapkan dapat dimanfaatkan oleh instansi pemerintah maupun pengembang sistem publik untuk memantau persepsi masyarakat secara otomatis, dengan rekomendasi penelitian selanjutnya mencakup eksplorasi model multilingual dan perluasan data dari berbagai platform media sosial.

Kata kunci: Analisis Sentimen, Coretax, *IndoBERT*, *IndoBERTweet*, *IndoRoBERTa*

ABSTRACT

This study aims to compare the performance of three Indonesian monolingual Transformer models, namely *IndoBERT*, *IndoBERTweet*, and *IndoRoBERTa*, in sentiment analysis of public opinions regarding the Coretax website on platform X. The research process involved data collection, text preprocessing, data augmentation, and fine-tuning of each model. Performance evaluation was conducted using the *classification report* based on precision, recall, f1-score, and accuracy metrics. The results indicate that *IndoBERT* achieved the best performance compared to the other two models, with the highest macro-average scores across all evaluation metrics. These findings are expected to support government institutions and public system developers in automatically monitoring public perception, with future research recommendations including the exploration of multilingual models and the expansion of data from various social media platforms.

Keyword: Sentiment Analysis, Coretax, IndoBERT, IndoBERTweet, IndoRoBERTa