

**ANALISIS SENTIMEN TEKNOLOGI AI PADA MEDIA
SOSIAL X (TWITTER) MENGGUNAKAN MODEL
TRANSFORMER DENGAN PENERAPAN TEKNIK
*BALANCING***

SKRIPSI

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



disusun oleh

DYAH PRAMUDYA SARI

22.11.5097

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS SENTIMEN TEKNOLOGI AI PADA MEDIA
SOSIAL X (TWITTER) MENGGUNAKAN MODEL
TRANSFORMER DENGAN PENERAPAN TEKNIK
*BALANCING***

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DYAH PRAMUDYA SARI

22.11.5097

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN TEKNOLOGI AI PADA MEDIA
SOSIAL X (TWITTER) MENGGUNAKAN MODEL
TRANSFORMER DENGAN PENERAPAN TEKNIK
BALANCING**

yang disusun dan diajukan oleh

Dyah Pramudya Sari

22.11.5097

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Mei 2026

Dosen Pembimbing,



Rumini, S.Kom., M.Kom
NIK. 190302246

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS SENTIMEN TEKNOLOGI AI PADA MEDIA SOSIAL X
(TWITTER) MENGGUNAKAN MODEL *TRANSFORMER* DENGAN
PENERAPAN TEKNIK *BALANCING*

yang disusun dan diajukan oleh

Dyah Pramudya Sari

22.11.5097

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya Dhuhita, S.Kom., M.Kom.
NIK. 190302185

Andriyan Dwi Putra, S.Kom., M.Kom.
NIK. 190302270

Rumini, S.Kom., M.Kom.
NIK. 190302246



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Dyah Pramudya Sari
NIM : 22.11.5097

Menyatakan bahwa Skripsi dengan judul berikut:

**Analisis Sentimen Teknologi AI Pada Media Sosial X (Twitter)
Menggunakan Model *Transformer* Dengan Penerapan Teknik *Balancing***

Dosen Pembimbing : Rumini, 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, 22 Mei 2026

Yang Menyatakan,


Dyah Pramudya Sari

HALAMAN PERSEMBAHAN

“Jika Bukan Karena Allah yang Memampukan, Aku Mungkin Sudah Lama
Menyerah”

Alhamdulillahirobbil Alamin, Puji dan syukur penulis panjatkan kepada Allah SWT yang maha pengasih dan maha pemurah atas segala rahmat dan karunia yang telah dilimpahkan-Nya sehingga skripsi ini dapat diselesaikan dengan baik. Sungguh perjuangan yang cukup panjang telah di lalui untuk mendapatkan gelar sarjana ini. Dengah segala kerendahan hati dan rasa syukur yang mendalam, skripsi ini penulis persembahkan kepada:

1. Cinta pertama, Ayahanda Hardi tercinta, terima kasih selalu mengusahakan yang terbaik, terima kasih atas setiap cucuran keringat dan kerja keras yang telah engkau tukarkan menjadi sebuah naskah demi penulis dapat mengenyam pendidikan sampai di tingkat ini, memberikan dukungan, motivasi, kepercayaan serta selalu mengajarkan kebaikan dalam hidup penulis.
2. Pintu surga, Ibunda Jumiah tercinta, yang telah membesarkan penulis dengan penuh kasih sayang tanpa batas dan pengorbanan yang tidak terhitung, terimakasih atas segala doa, pesan, dan harapan yang selalu mendampingi setiap langkah dan usaha penulis untuk menjadi seseorang yang berpendidikan.
3. Seseorang yang terkasih dan tidak kalah penting kehadirannya, Ikhwan Nurramadhan yang memberikan semangat, dukungan serta bantuan baik itu tenaga, pikiran dan waktu sehingga penulis dapat menyelesaikan skripsi ini. Terimakasih telah menjadi bagian dari perjalanan ini, menjadi rumah tempat berkeluh kesah di waktu lelah, menjadi pendengar yang baik serta senantiasa memberikan cinta dan perhatian kepada penulis.

4. Keluarga besar terima kasih atas doa, dukungan dan cinta yang senantiasa menguatkan di setiap langkah perjuangan ini. Terima kasih atas kehangatan dan ketulusannya.
5. Alm Kakek dan Almh Nenek, meskipun raga mu telah tiada, doa-doa dan harapanmu hidup dalam langkah ini.
6. Sahabat-sahabat terkasih, berjumpa di bangku perkuliahan adalah anugrah yang tak pernah penulis duga. Terima kasih telah berjalan bersama, terima kasih atas segala motivasi, dukungan, pengalaman, waktu dan ilmu selama masa perkuliahan.
7. Universitas Amikom Yogyakarta, almamater yang telah menjadi tempat penulis belajar, berkembang, dan tumbuh menjadi pribadi yang lebih dewasa. Terima kasih atas pengalaman, ilmu dan kesempatan yang telah diberikan.

Semoga segala usaha, doa, dukungan, dan pengorbanan yang telah diberikan kepada penulis mendapatkan balasan terbaik dari Allah SWT. Penulis menyadari bahwa skripsi ini masih memiliki banyak kekurangan, namun besar harapan agar karya ini dapat memberikan manfaat, menambah wawasan, serta menjadi langkah awal bagi penulis untuk terus belajar, berkembang, dan memberikan kontribusi yang baik di masa mendatang.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT atas segala rahmat, hidayah serta karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Analisis Sentimen Teknologi AI Pada Media Sosial X (Twitter) Menggunakan Model Transformer Dengan Penerapan Teknik *Balancing*"** dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Penulis menyadari bahwa penyusunan skripsi ini masih terdapat berbagai kekurangan yang dipengaruhi oleh keterbatasan pengetahuan, pengalaman, serta kemampuan yang dimiliki penulis. Namun berkat doa, dukungan, bantuan, serta bimbingan dari berbagai pihak, skripsi ini akhirnya dapat diselesaikan dengan baik. Pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih dan penghargaan yang tulus 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 Universitas Amikom Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom, selaku Ketua Program Studi Informatika Universitas Amikom Yogyakarta.
4. Ibu Rumini, M.Kom, selaku Dosen Wali dan Dosen Pembimbing yang dengan penuh kesabaran dan perhatian telah meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis hingga sampai pada titik ini. Terima kasih selalu memberikan bantuan, arahan dan mendorong penulis untuk terus berkembang.
5. Seluruh Dosen dan Staf di Universitas Amikom Yogyakarta, yang telah dengan tulus membagikan ilmu, pengalaman, arahan dan bantuan selama penulis menempuh pendidikan

6. Kedua orang tua yang selalu menjadi tempat pulang, sumber kekuatan dan alasan terbesar penulis untuk terus bertahan dan berjuang. Terima kasih atas setiap doa yang tidak pernah putus, pengorbanan yang tidak pernah diperhitungkan serta kasih sayang yang selalu mengiringi langkah penulis dalam keadaan apa pun.
7. Seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah hadir dan memberikan dukungan, bantuan, semangat serta doa selama proses penyusunan skripsi ini.

Penulis berharap skripsi ini dapat memberikan manfaat serta menambah wawasan bagi para pembaca dan pihak-pihak yang membutuhkan. Penulis juga menyadari bahwa karya ini masih memiliki berbagai keterbatasan, sehingga kritik dan saran yang membangun sangat diharapkan. Semoga segala bantuan, dukungan, dan kebaikan yang telah diberikan kepada penulis mendapatkan balasan yang terbaik dari Allah SWT.

Yogyakarta, 20 Mei 2026
Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xvi
INTISARI	xxv
<i>ABSTRACT</i>	xxvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
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	14
BAB III METODE PENELITIAN		23
3.1	Objek Penelitian	23
3.2	Alur Penelitian	23
3.3	Alat dan Bahan	32
BAB IV HASIL DAN PEMBAHASAN		34
4.1.	Pengumpulan Data	34
4.2	Pelabelan Data	35
4.3	<i>Preprocessing Data</i>	36
4.4	<i>Splitting Data</i>	39
4.5	Implementasi Balancing Data dan Tokenisasi	41
4.6	Hasil Pelatihan dan Evaluasi Model (<i>Fine-Tuning</i>)	43
4.7	Komparasi dan Analisis Kinerja Model Terbaik	48
4.8	Implementasi dan Pengujian Model pada Sistem <i>Deploy</i>	51
BAB V PENUTUP		54
5.1	Kesimpulan	54
5.2	Saran	55
REFERENSI		57
LAMPIRAN		62

DAFTAR TABEL

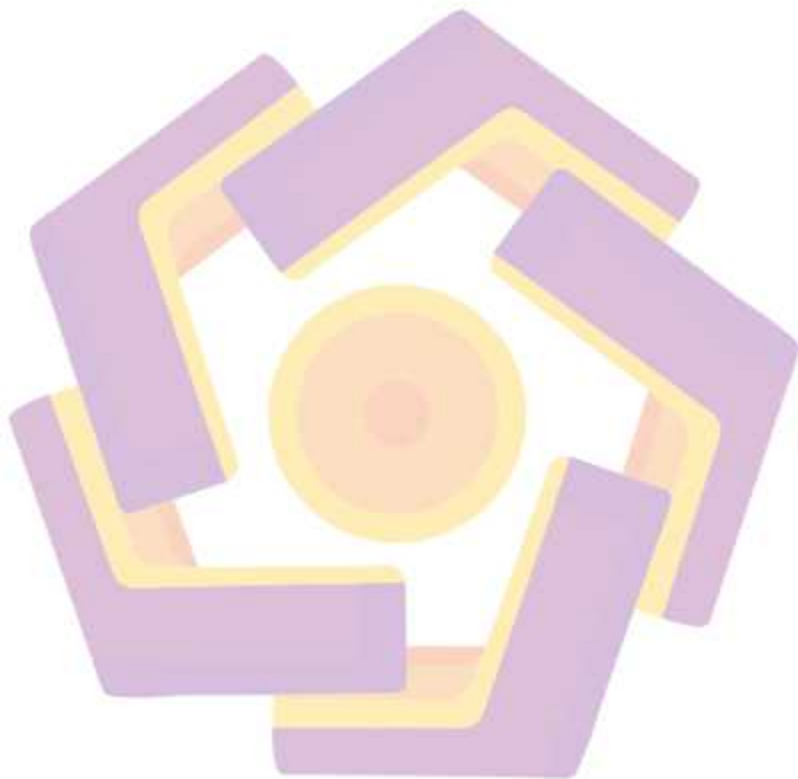
Tabel 2.1 Keaslian Penelitian	9
Tabel 2.2 <i>Confusion Matrix</i>	22
Tabel 3.1 Perhitungan <i>Softmax</i> Data Positif	30
Tabel 3.2 Perhitungan <i>Softmax</i> Data Negatif.....	31
Tabel 4.1 Contoh Hasil Transformasi Teks pada Tahap <i>Preprocessing</i>	39
Tabel 4.2 Distribusi Data Bersih (<i>Clean Dataset</i>) Berdasarkan Label Sentimen..	39
Tabel 4.3 Distribusi Kelas Sentimen pada Data Latih dan Data Validasi	41
Tabel 4.4 Laporan Klasifikasi (<i>Classification Report</i>) Skenario <i>Baseline</i>	44
Tabel 4.5 Laporan Klasifikasi (<i>Classification Report</i>) Skenario ROS	46
Tabel 4.6 Laporan Klasifikasi (<i>Classification Report</i>) Skenario RUS	47
Tabel 4.7 Laporan Klasifikasi (<i>Classification Report</i>) Skenario <i>Class Weight</i>	48
Tabel 4.8 Komparasi Kinerja Keseluruhan Skenario Eksperimen IndoBERT	49

DAFTAR GAMBAR

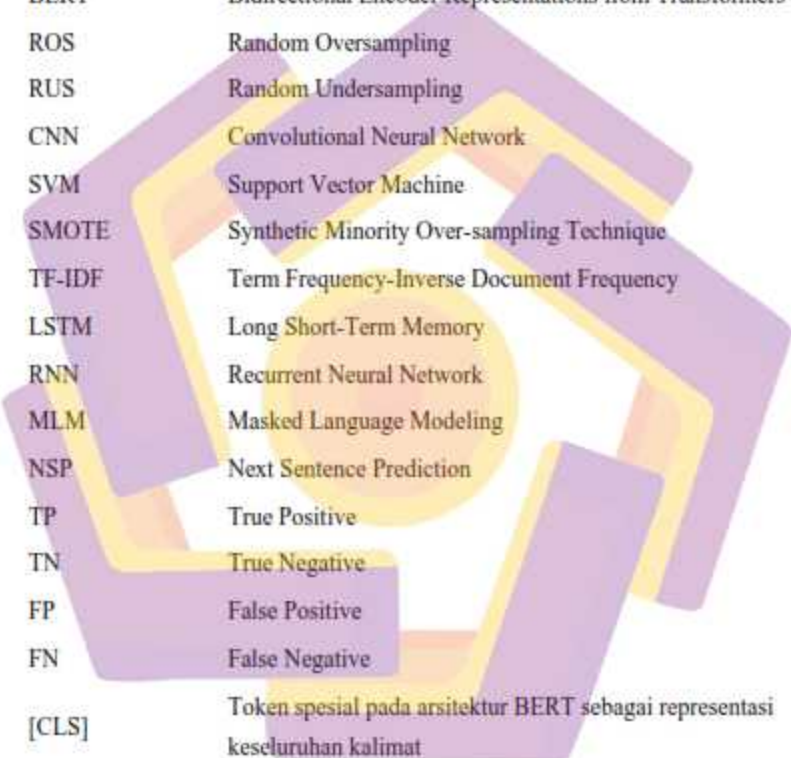
Gambar 2.1 Diagram Arsitektur <i>Fine-Tuning</i> IndoBERT	18
Gambar 3.1 Alur Penelitian	25
Gambar 4.1 Cuplikan Data Mentah (<i>Raw Data</i>) Hasil <i>Crawling</i> Menggunakan <i>Tweet-Harvest</i>	36
Gambar 4.2 Cuplikan Data Terintegrasi Label Sentimen	37
Gambar 4.3 Perbandingan Distribusi Kelas Sentimen pada Berbagai Skenario Data Latih	42
Gambar 4.4 Cuplikan Representasi Numerik Tensor Hasil Tokenisasi IndoBERT	43
Gambar 4.5 <i>Confusion Matrix</i> Model IndoBERT - Skenario <i>Baseline</i>	44
Gambar 4.6 <i>Confusion Matrix</i> Model IndoBERT - Skenario <i>ROS</i>	45
Gambar 4.7 <i>Confusion Matrix</i> Model IndoBERT - Skenario <i>RUS</i>	46
Gambar 4.8 <i>Confusion Matrix</i> Model IndoBERT - Skenario <i>Class Weight</i>	48
Gambar 4.9 Hasil Prediksi Sentimen Negatif pada Sistem <i>Deploy</i>	52
Gambar 4.10 Hasil Prediksi Sentimen Positif pada Sistem <i>Deploy</i>	53

DAFTAR LAMPIRAN

Lampiran 1. Dokumentasi Ekstraksi Data Dengan <i>Tweet Harvest</i>	62
Lampiran 2. <i>Listing Program (Souce Code Python)</i>	62



DAFTAR LAMBANG DAN SINGKATAN



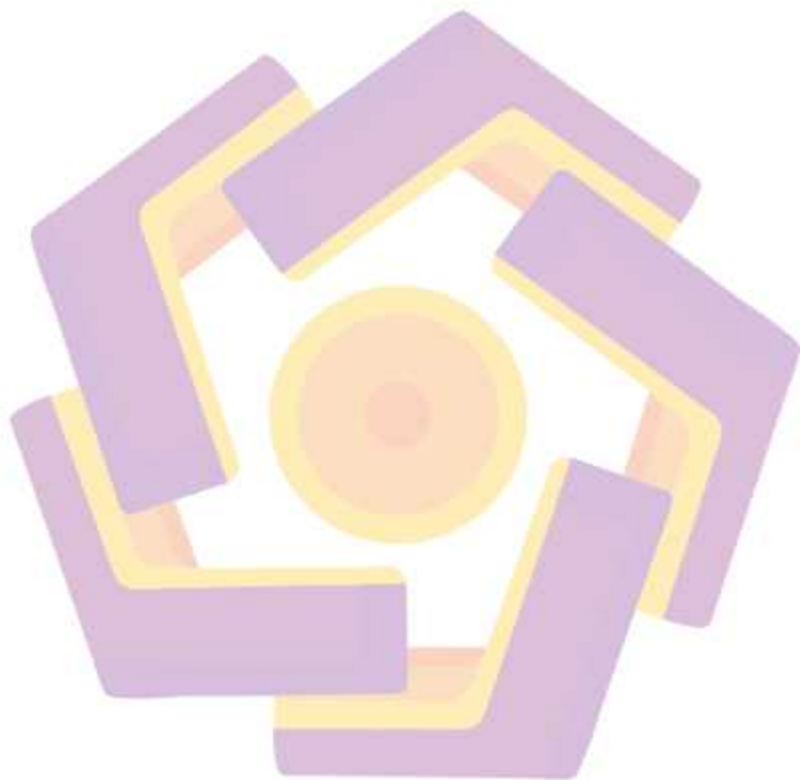
AI	Artificial Intelligence
NLP	Natural Language Processing
X	Platform media sosial yang sebelumnya Twitter
BERT	Bidirectional Encoder Representations from Transformers
ROS	Random Oversampling
RUS	Random Undersampling
CNN	Convolutional Neural Network
SVM	Support Vector Machine
SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency-Inverse Document Frequency
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
MLM	Masked Language Modeling
NSP	Next Sentence Prediction
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
[CLS]	Token spesial pada arsitektur BERT sebagai representasi keseluruhan kalimat
σ	Fungsi aktivasi softmax
z	Vektor output (logits) model
e	Bilangan Euler
K	Jumlah total kelas klasifikasi
GPU	Graphics Processing Unit

CSV

Comma Separated Values

NaN

Not a Number



DAFTAR ISTILAH



Analisis Sentimen	Teknik analisis teks untuk mengidentifikasi opini atau emosi dalam suatu data teks.
AI Anxiety	Kecemasan terhadap perkembangan atau penggunaan Artificial Intelligence
Media Sosial X	Platform media sosial berbasis microblogging yang digunakan sebagai sumber data penelitian.
Transformer	Arsitektur deep learning yang mampu memahami hubungan konteks dalam teks menggunakan mekanisme attention.
IndoBERT	Model transformer berbasis BERT yang dilatih khusus menggunakan korpus bahasa Indonesia.
Imbalanced Data	Kondisi distribusi data yang tidak seimbang antar kelas.
Balancing Data	Teknik untuk menyeimbangkan distribusi jumlah data antar kelas.
Random Oversampling	Teknik balancing dengan menambah jumlah data pada kelas minoritas secara acak.
Random Undersampling	Teknik balancing dengan mengurangi jumlah data pada kelas mayoritas secara acak.
Class Weight	Teknik penyesuaian bobot kelas pada proses pelatihan model untuk mengurangi bias klasifikasi.
Preprocessing	Tahapan pembersihan dan normalisasi data teks sebelum diproses model.
Model Pre-trained	Model yang telah dilatih sebelumnya menggunakan data dalam jumlah besar.
Evaluasi Model	Proses pengukuran performa model klasifikasi



Dataset	Kumpulan data yang digunakan dalam penelitian.
Convolutional Neural Network	Arsitektur deep learning yang digunakan untuk ekstraksi fitur dan klasifikasi data.
Naïve Bayes	Algoritma klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi independensi fitur.
Support Vector Machine	Algoritma machine learning yang bekerja dengan mencari hyperplane pemisah terbaik antar kelas.
SMOTE	Teknik balancing dengan membuat sampel sintetis pada kelas minoritas.
Fine-Tuning	Proses penyesuaian model pre-trained agar optimal pada tugas tertentu.
Deep Learning	Cabang machine learning berbasis jaringan saraf berlapis untuk mempelajari pola kompleks pada data.
Random Forest	Algoritma ensemble learning berbasis kumpulan decision tree untuk meningkatkan stabilitas klasifikasi.
Machine Learning	Metode kecerdasan buatan yang memungkinkan sistem belajar dari data secara otomatis.
Confusion Matrix	Tabel evaluasi klasifikasi yang membandingkan hasil prediksi dengan label sebenarnya.
Grid Search	Teknik pencarian kombinasi hyperparameter terbaik secara sistematis.
Dropout	Teknik regularisasi dengan menonaktifkan neuron secara acak selama pelatihan model.
Learning Rate	Parameter yang menentukan besar langkah pembaruan bobot model selama proses training.

AI Self-Efficacy	Tingkat keyakinan individu terhadap kemampuannya dalam memahami dan menggunakan AI.
Microblogging	Bentuk media sosial berbasis unggahan teks singkat.
Hashtag	Penanda kata kunci menggunakan simbol # untuk mengelompokkan topik tertentu.
Trending Topics	Topik yang sedang populer dan banyak dibahas pada media sosial.
Opinion Mining	Istilah lain dari analisis sentimen yang berfokus pada identifikasi opini dalam teks.
Lexicon-Based	Pendekatan analisis sentimen berbasis kamus kata yang memiliki nilai sentimen tertentu.
Bag-of-Words	Representasi teks berdasarkan frekuensi kata tanpa mempertimbangkan urutan kata.
Word Embedding	Representasi kata dalam bentuk vektor numerik berdasarkan konteks semantik.
Case Folding	Proses mengubah seluruh huruf menjadi lowercase untuk konsistensi data.
Cleaning	Proses membersihkan noise seperti simbol, URL, atau karakter tidak relevan dari teks.
Normalisasi	Proses mengubah kata tidak baku menjadi bentuk baku.
Self-Attention	Mekanisme pada Transformer untuk menentukan hubungan penting antar kata dalam kalimat.
Long-Range Dependencies	Kemampuan model memahami hubungan antar kata yang berjauhan dalam teks.
Bidirectional	Pemrosesan konteks teks dari dua arah, kiri dan kanan.



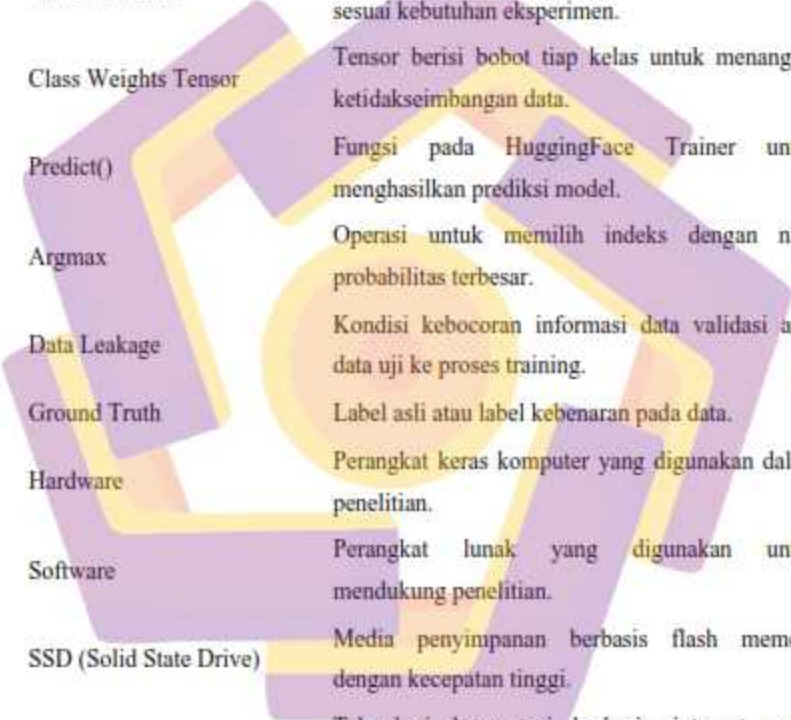
Pre-Training Objectives	Tugas awal pelatihan model bahasa sebelum digunakan pada tugas spesifik.
Masked Language Modeling	Metode pelatihan model dengan memprediksi kata yang disembunyikan.
Next Sentence Prediction	Metode pelatihan model untuk memahami hubungan antar kalimat.
Pooler Output	Representasi akhir keluaran model yang mewakili keseluruhan konteks kalimat.
Dense Layer	Lapisan neural network yang seluruh neuronnya saling terhubung penuh.
Softmax	Fungsi aktivasi yang mengubah logits menjadi distribusi probabilitas.
Logits	Nilai output mentah model sebelum diproses fungsi aktivasi.
Kelas Mayoritas	Kelas dengan jumlah data paling banyak dalam dataset.
Kelas Minoritas	Kelas dengan jumlah data lebih sedikit dalam dataset.
Real-World Data	Data nyata yang diperoleh langsung dari lingkungan atau aktivitas sebenarnya.
Data-Level	Pendekatan penanganan imbalance melalui manipulasi data.
Algorithmic-Level	Pendekatan penanganan imbalance melalui modifikasi algoritma pembelajaran.
Sampling With Replacement	Teknik pengambilan sampel yang memungkinkan data terpilih lebih dari satu kali.
Overfitting	Kondisi model terlalu menghafal data latih sehingga performanya buruk pada data baru.



Training	Proses pelatihan model menggunakan data latih.
Sampling Without Replacement	Teknik pengambilan sampel tanpa pengulangan data yang sudah dipilih.
Information Loss	Hilangnya informasi penting akibat pengurangan data.
Decision Boundary	Batas keputusan model dalam memisahkan kelas data.
Cost-Sensitive Learning	Pendekatan pembelajaran yang memberikan bobot kesalahan berbeda pada tiap kelas.
Weight Penalty	Bobot penalti yang diberikan pada fungsi loss model.
Loss Function	Fungsi matematis untuk mengukur kesalahan prediksi model.
Cross-Entropy Loss	Fungsi loss untuk mengukur kesalahan klasifikasi berdasarkan probabilitas prediksi.
Accuracy	Metrik evaluasi yang mengukur rasio prediksi benar terhadap seluruh data.
Precision	Metrik evaluasi yang mengukur ketepatan prediksi positif model.
Recall	Metrik evaluasi yang mengukur kemampuan model menemukan seluruh data positif.
F1-Score	Rata-rata harmonis antara precision dan recall.
Crawling Data	Proses pengambilan data secara otomatis dari media sosial atau website menggunakan program tertentu.
Auth Token	Token autentikasi yang digunakan untuk mengakses layanan atau akun tertentu secara aman.

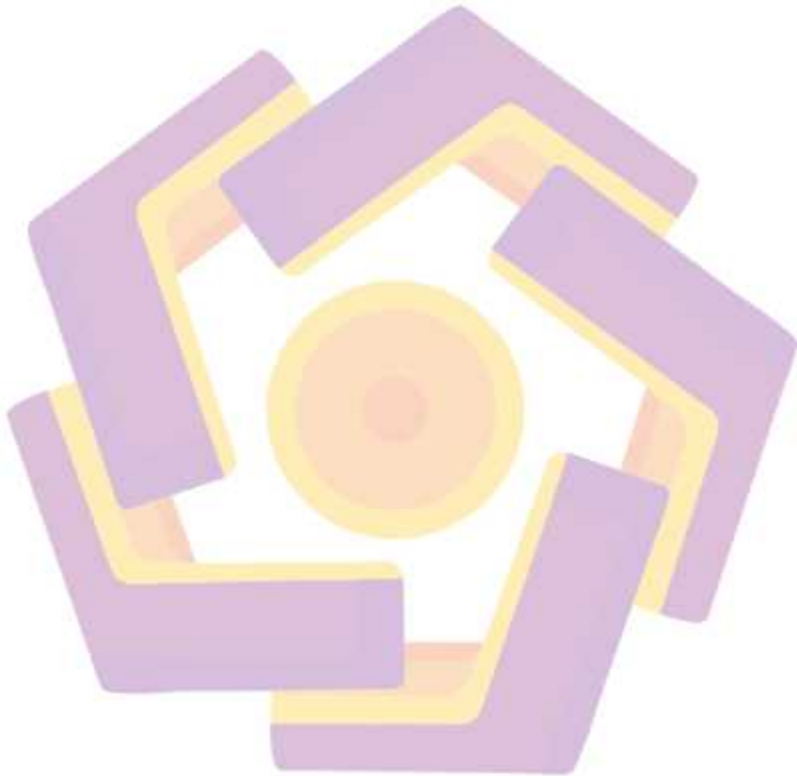
Tweet-Harvest	Tools berbasis Python yang digunakan untuk melakukan crawling data dari media sosial X.
Google Colaboratory (Google Colab)	Platform cloud berbasis notebook untuk menjalankan kode Python secara online.
Query	Kata kunci atau kombinasi kata yang digunakan untuk mengambil data tertentu.
Shuffling	Proses pengacakan urutan data agar distribusi data lebih acak dan tidak berpola.
Representatif	Kondisi data yang dianggap mampu mewakili karakteristik populasi sebenarnya.
Noise	Data atau karakter yang tidak relevan dan dapat mengganggu proses analisis.
Lowercase	Format huruf kecil pada teks hasil preprocessing.
Mention	Penanda akun pengguna pada media sosial menggunakan simbol @.
URL	Alamat atau tautan suatu halaman web di internet.
Dictionary-Based	Pendekatan berbasis kamus pasangan kata untuk proses normalisasi teks.
Stratified Sampling	Teknik pembagian data dengan mempertahankan proporsi kelas pada setiap subset data.
Training Data	Data yang digunakan untuk melatih model machine learning.
Validation Data	Data yang digunakan untuk mengevaluasi performa model selama pelatihan.
Random State	Parameter untuk menjaga konsistensi hasil pengacakan data agar dapat direproduksi.
Tokenisasi	Proses mengubah teks menjadi unit-unit token yang dapat diproses model NLP.

Token	Unit teks berupa kata atau subword hasil proses tokenisasi.
Subword Token	Potongan kata yang digunakan model Transformer untuk merepresentasikan teks.
Padding	Proses penambahan token kosong agar panjang input seragam.
Truncation	Proses pemotongan teks yang melebihi panjang maksimum token.
Input IDs	Representasi numerik token berdasarkan indeks kosakata model.
Attention Mask	Penanda token asli dan token padding pada input model Transformer.
Token Type IDs	Penanda segmen kalimat pada input model BERT.
Tensor	Struktur data multidimensi yang digunakan dalam komputasi deep learning.
PyTorch	Framework deep learning berbasis Python untuk pengembangan model AI.
BertForSequenceClassification	Modul HuggingFace Transformers untuk tugas klasifikasi berbasis BERT.
Probabilitas	Nilai kemungkinan kemunculan suatu kelas hasil prediksi model.
Epoch	Satu siklus penuh pelatihan model terhadap seluruh data latih.
Batch Size	Jumlah data yang diproses model dalam satu iterasi pelatihan.
Warmup Steps	Tahapan awal pelatihan untuk menyesuaikan learning rate secara bertahap.
Weight Decay	Teknik regularisasi untuk mengurangi overfitting



Hyperparameter	dengan penalti pada bobot model. Parameter konfigurasi model yang ditentukan sebelum proses training dimulai.
HuggingFace Transformers	Library NLP berbasis Transformer untuk implementasi model bahasa modern.
Custom Trainer	Pelatih model yang dimodifikasi secara khusus sesuai kebutuhan eksperimen.
Class Weights Tensor	Tensor berisi bobot tiap kelas untuk menangani ketidakseimbangan data.
Predict()	Fungsi pada HuggingFace Trainer untuk menghasilkan prediksi model.
Argmax	Operasi untuk memilih indeks dengan nilai probabilitas terbesar.
Data Leakage	Kondisi kebocoran informasi data validasi atau data uji ke proses training.
Ground Truth	Label asli atau label kebenaran pada data.
Hardware	Perangkat keras komputer yang digunakan dalam penelitian.
Software	Perangkat lunak yang digunakan untuk mendukung penelitian.
SSD (Solid State Drive)	Media penyimpanan berbasis flash memory dengan kecepatan tinggi.
Cloud Computing	Teknologi komputasi berbasis internet untuk menjalankan proses pengolahan data secara daring.
Feature Selection	Proses pemilihan atribut atau fitur yang relevan untuk digunakan dalam pemodelan.
Classification Report	Laporan evaluasi model klasifikasi yang memuat

Confidence Score	metrik seperti precision, recall, dan F1-score. Tingkat keyakinan model terhadap hasil prediksi suatu kelas.
Deploy / Deployment	Proses implementasi model ke dalam sistem atau aplikasi agar dapat digunakan secara langsung.



INTISARI

Perkembangan teknologi kecerdasan buatan (AI) yang semakin pesat telah mendorong transformasi digital di berbagai sektor dan membawa perubahan signifikan dalam berbagai aktivitas manusia. Namun di sisi lain, perkembangan tersebut juga memunculkan beragam respons dan persepsi masyarakat terkait dampak penggunaan AI. Berbagai opini banyak diekspresikan melalui media sosial X dalam bentuk unggahan dan komentar. Banyaknya data teks pada media sosial menyebabkan analisis manual menjadi kurang efektif, sehingga diperlukan pendekatan analisis sentimen berbasis komputasional untuk mengidentifikasi persepsi masyarakat terhadap perkembangan AI secara otomatis.

Penelitian ini menerapkan metode analisis sentimen berbasis model Transformer IndoBERT untuk mengklasifikasikan sentiment teknologi AI ke dalam kelas positif dan negatif. Data penelitian diperoleh melalui proses *crawling* pada media sosial X dengan total dataset sebanyak 4.689 data. Tahapan penelitian meliputi *preprocessing* teks, *balancing* data menggunakan metode *Random Oversampling* (ROS), *Random Undersampling* (RUS), dan *Class Weight*, proses *fine-tuning* model, serta evaluasi menggunakan *confusion matrix* dan *classification report*.

Hasil penelitian menunjukkan bahwa sentimen negatif terhadap AI lebih dominan dibandingkan sentimen positif. Model IndoBERT berhasil mencapai performa sangat baik dengan nilai akurasi dan rata-rata *F1-Score* sebesar 0,99 pada seluruh skenario eksperimen. Pendekatan *Class Weight* menghasilkan performa paling optimal dengan nilai *validation loss* terendah dan jumlah misklasifikasi paling sedikit. Penelitian ini diharapkan dapat menjadi referensi pengembangan sistem analisis sentimen berbasis AI untuk memahami persepsi masyarakat terhadap perkembangan teknologi AI melalui media sosial.

Kata kunci: Teknologi AI, Analisis Sentimen, IndoBERT, Transformer, Class Weight.

ABSTRACT

The rapid development of Artificial Intelligence (AI) technology has driven digital transformation across various sectors and brought significant changes to many aspects of human activities. On the other hand, these developments have also generated various public responses and perceptions regarding the impact of AI usage. Many of these opinions are expressed through social media platform X in the form of posts and comments. The large volume of textual data on social media makes manual analysis less effective, thus requiring a computational sentiment analysis approach to automatically identify public perceptions toward AI development.

This study implements a sentiment analysis method based on the IndoBERT Transformer model to classify sentiments toward AI technology into positive and negative classes. The dataset was collected through a crawling process on social media platform X, resulting in a total of 4,689 data samples. The research stages include text preprocessing, data balancing using Random Oversampling (ROS), Random Undersampling (RUS), and Class Weight methods, model fine-tuning, and evaluation using confusion matrix and classification report.

The results indicate that negative sentiment toward AI is more dominant than positive sentiment. The IndoBERT model achieved excellent performance with an accuracy score and average F1-Score of 0.99 across all experimental scenarios. The Class Weight approach produced the most optimal performance with the lowest validation loss and the fewest misclassifications. This research is expected to serve as a reference for developing AI-based sentiment analysis systems to better understand public perceptions of AI technology development through social media.

Keyword: *AI Technology, Sentiment Analysis, IndoBERT, Transformer, Class Weight.*