

**ANALISIS SENTIMEN DAN UJARAN KEBENCIAN PADA KOMENTAR
YOUTUBE VIDEO NAJWA SHIHAB ‘SUSAHNYA JADI PEREMPUAN’
MENGUNAKAN INDOBERT**

SKRIPSI

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



disusun oleh

IKA SAFITRI

22.11.5153

Kepada

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

**ANALISIS SENTIMEN DAN UJARAN KEBENCIAN PADA KOMENTAR
YOUTUBE VIDEO NAJWA SHIHAB ‘SUSAHNYA JADI PEREMPUAN’
MENGUNAKAN INDOBERT**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

IKA SAFITRI

22.11.5153

Kepada

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

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN DAN UJARAN KEBENCIAN PADA KOMENTAR
YOUTUBE VIDEO NAJWA SHIHAB ‘SUSAHNYA JADI PEREMPUAN’
MENGUNAKAN INDOBERT**

yang disusun dan diajukan oleh

Ika Safitri

22.11.5153

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 23 Februari 2026

Dosen Pembimbing,



Bayu Setiaji, M.Kom.

NIK. 190302216

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN DAN UJARAN KEBENCIAN PADA KOMENTAR
YOUTUBE VIDEO NAJWA SHIHAB ‘SUSAHNYA JADI PEREMPUAN’
MENGUNAKAN INDOBERT**

yang disusun dan diajukan oleh

Ika Safitri
22.11.5153

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Februari 2026

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
<u>Bambang Piliu Hartato, S.Kom., M.Eng</u> NIK. 190302707	
<u>Ali Mustopa, S.Kom., M.Kom</u> NIK. 190302192	
<u>Bayu Setiaji, M.Kom</u> NIK. 190302216	

Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ika Safitri
NIM : 22.11.5153

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Sentimen dan Ujaran Kebencian pada Komentar YouTube Video Najwa Shihab ‘Susahnya Jadi Perempuan’ Menggunakan IndoBERT.

Dosen Pembimbing : Bayu Setiaji, 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 Februari 2026

Yang Menyatakan,

Meterai Asli
Rp 10.000,-



Ika Safitri

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan dengan penuh rasa syukur dan terima kasih kepada:

1. Mama tercinta, yang selalu mendoakan, mendukung, dan menguatkan penulis tanpa henti. Terima kasih atas kesabaran, perhatian, dan kehadiran yang selalu mendengarkan keluh kesah penulis serta memberi ketenangan dalam setiap proses yang dilalui.
2. Ayah tercinta, atas doa, kerja keras, kesabaran, dan bimbingan yang selalu diberikan kepada penulis. Terima kasih atas arahan, keteguhan, dan perjuangan yang terus mengutamakan keluarga, serta menjadi sosok yang sangat penulis hormati dan kagumi.
3. Bibi tersayang, yang telah memberikan kasih sayang, doa, perhatian, dan dukungan yang tulus selama ini. Terima kasih atas kebaikan, kepedulian, dan bantuan yang diberikan kepada penulis, serta rasa sayang yang dirasakan penulis dengan begitu nyata.
4. Kakak dan adik penulis, dengan terima kasih khusus kepada kakak penulis atas dukungan, pengertian, dan motivasi yang sangat membantu, terutama pada awal dimulainya skripsi ini. Untuk adik-adik penulis yang selalu disayangi, terima kasih atas doa dan semangatnya.
5. Teman-teman seperjuangan, yang telah menemani proses penyusunan skripsi, menjadi tempat berbagi cerita, saling menyemangati, dan berjalan bersama dalam masa perjuangan ini.
6. Terakhir, untuk diri penulis sendiri, terima kasih telah bertahan, berjuang, dan terus mencoba hingga sampai pada tahap ini. Perjalanan ini tidak selalu mudah dan masih banyak hal yang perlu dipelajari, namun setiap usaha dan keteguhan untuk tidak berhenti telah membawa penulis sejauh ini. Terima kasih telah tetap berusaha kuat di tengah kelelahan dan tekanan. Semoga ke depan penulis terus bertumbuh, berani mencoba hal baru, dan melangkah menjadi pribadi yang lebih baik.

KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas rahmat dan karunia-Nya sehingga skripsi berjudul “Analisis Sentimen dan Deteksi Ujaran Kebencian pada Komentar YouTube Video Najwa Shihab ‘Susahnya Jadi Perempuan’ Menggunakan IndoBERT” dapat diselesaikan sebagai salah satu syarat memperoleh gelar Sarjana. Dalam proses penyusunan skripsi ini, penulis memperoleh banyak arahan dan dukungan dari berbagai pihak. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Prof. Dr. Mohammad Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta, atas penyelenggaraan pendidikan, dukungan lingkungan akademik, serta fasilitas yang diberikan selama penulis menempuh proses perkuliahan hingga tahap penyusunan skripsi.
2. Bapak Bayu Setiaji M.Kom., selaku dosen pembimbing, yang telah memberikan bimbingan sejak tahap pengajuan judul hingga penyelesaian akhir skripsi ini. Terima kasih atas arahan, nasihat, dan masukan yang konsisten, serta kesediaan waktu dalam membimbing setiap tahap penelitian sehingga penulis dapat menyusun dan menyelesaikan skripsi ini dengan lebih terarah.
3. Orang tua penulis, atas doa, dukungan, kesabaran, dan pengorbanan yang tidak pernah terputus selama masa studi, serta atas kepercayaan dan dorongan yang selalu menguatkan penulis dalam menjalani proses perkuliahan hingga penyelesaian skripsi ini.
4. Keluarga penulis, termasuk kakak, adik, bibi, dan keluarga besar, yang telah memberikan dukungan serta doa bagi kelancaran penyusunan skripsi ini.

Yogyakarta, 12 Februari 2026

Penulis

DAFTAR ISI

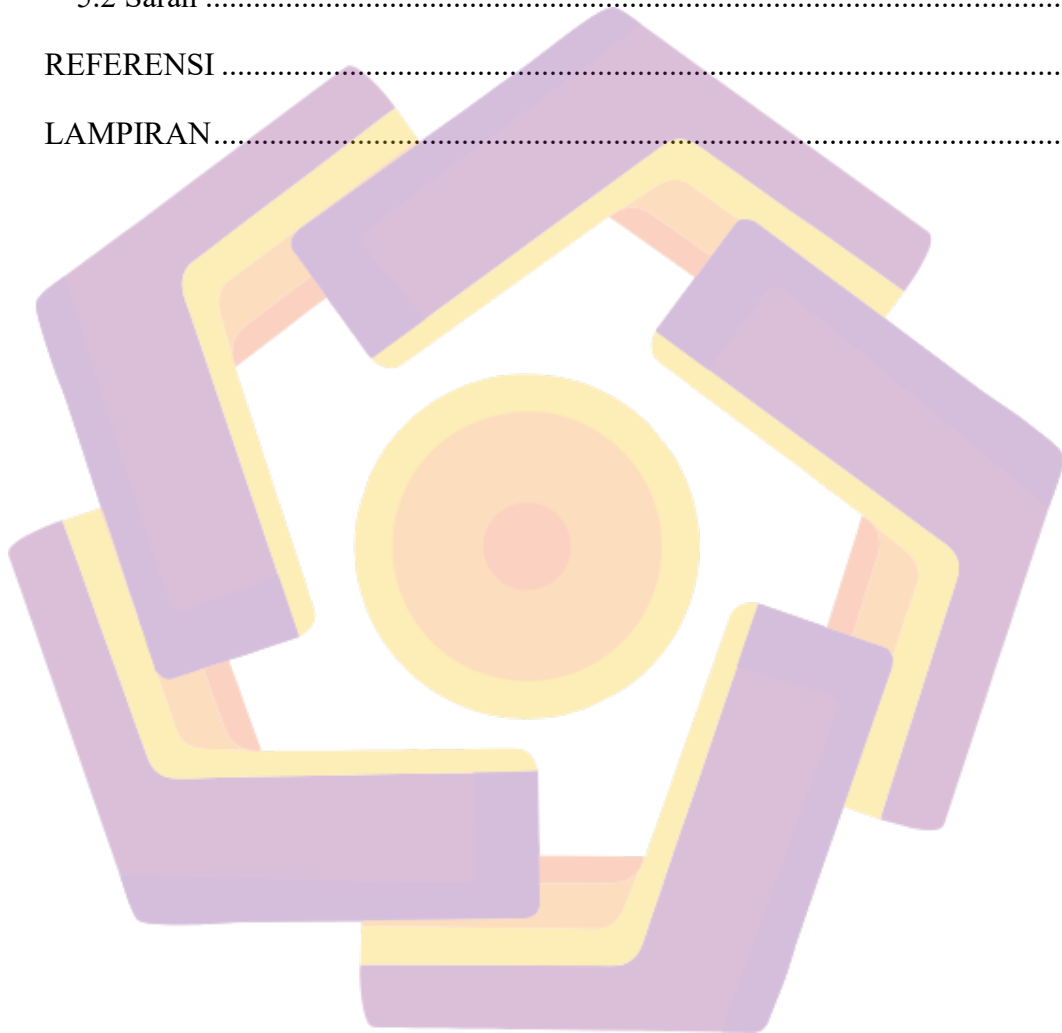
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
DAFTAR LAMBANG DAN SINGKATAN	xv
DAFTAR ISTILAH	xvi
INTISARI	xvii
<i>ABSTRACT</i>	xviii
BAB I.....	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II.....	5
TINJAUAN PUSTAKA	5

2.1 Studi Literatur	5
2.2 Dasar Teori.....	12
2.2.1 Natural Language Processing (NLP)	12
2.2.2 Text Processing	12
2.2.3 Model Transformer	13
2.2.4 Model BERT (Bidirectional Encoder Representations from Transformers).....	13
2.2.5 Model IndoBERT.....	14
2.2.6 Analisis Sentimen	15
2.2.7 Deteksi Ujaran Kebencian	16
2.2.8 Evaluasi Model	17
2.2.9 Sosial media YouTube.....	18
BAB III	20
METODE PENELITIAN.....	20
3.1 Objek Penelitian.....	20
3.2 Alur Penelitian	20
3.2.1 Pengumpulan Data	21
3.2.2 Pelabelan Data	21
3.2.3 Prapemrosesan Teks.....	23
3.2.4 Pembagian Data	24
3.2.5 Model Baseline	24
3.2.6 Pelatihan Model IndoBERT.....	25
3.2.6.1 Persiapan Input Model	25
3.2.6.2 Konfigurasi Pelatihan dan Hyperparameter	26
3.2.6.3 Penanganan Ketidakseimbangan Kelas dengan Class Weighting..	27

3.2.6.4 Pemilihan Model	27
3.2.7 Evaluasi Model	27
3.2.8 Analisis Hasil	28
3.3 Alat dan Bahan	28
3.3.1 Data Penelitian	28
3.3.2 Perangkat Keras	28
3.3.3 Perangkat Lunak	29
BAB IV	30
HASIL DAN PEMBAHASAN	30
4.1 Deskripsi Dataset dan Distribusi Data	30
4.1.1 Dataset Penelitian	30
4.1.2 Jumlah Data yang Digunakan	30
4.1.3 Panjang Komentar dan Word Cloud	32
4.1.4 Pembagian Data	34
4.2 Baseline Pemanding untuk Analisis Sentimen dan Deteksi Ujaran Kebencian	34
4.2.1 Baseline untuk Analisis Sentimen	34
4.2.1.1 Majority Class Baseline	34
4.2.1.2 TF-IDF dengan Logistic Regression	35
4.2.2 Baseline untuk Deteksi Ujaran Kebencian	36
4.2.2.1 Majority Class Baseline	36
4.2.2.2 Logistic Regression dengan Logistic Regression	36
4.3 Analisis Sentimen Menggunakan IndoBERT	37
4.3.1 Konfigurasi Model dan Proses Pelatihan	37
4.3.2 Penanganan Ketidakseimbangan Kelas dengan Class-Weighted	38

4.3.3 Hasil Pelatihan Model.....	38
4.3.4 Evaluasi Model pada Data Uji	39
4.3.5 Analisis Performa Klasifikasi per Kelas	40
4.3.6 Analisis Confusion Matrix	40
4.3.6.1 Kesalahan dominan antara kelas Negatif dan Netral	41
4.3.6.2 Kesalahan antara kelas Netral dan Negatif	42
4.3.6.3 Kesalahan antara kelas Positif dan Negatif.....	42
4.4 Deteksi Ujaran Kebencian Menggunakan IndoBERT	43
4.4.1 Konfigurasi Model dan Proses Pelatihan.....	43
4.4.2 Penanganan Ketidakseimbangan Kelas dengan Class-Weighted	43
4.4.3 Hasil Pelatihan Model.....	44
4.4.4 Evaluasi Model pada Data Uji	45
4.4.5 Analisis Performa Klasifikasi per Kelas	45
4.4.6 Analisis Confusion Matrix	46
4.4.6.1 Kesalahan False Positive.....	47
4.4.6.2 Kesalahan False Negative	47
4.5 Analisis Pola Opini Publik terhadap Isu Kesetaraan Gender.....	48
4.5.1 Distribusi Sentimen dan Karakteristiknya	48
4.5.2 Perbedaan antara Sentimen Negatif dan Ujaran Kebencian	50
4.5.3 Pola Opini dan Implikasinya.....	51
4.6 Perbandingan Model dan Pembahasan	52
4.6.1 Perbandingan Kinerja Model Baseline dan IndoBERT Analisis Sentimen	52
4.6.2 Perbandingan Kinerja Model Baseline dan IndoBERT Deteksi Ujaran Kebencian	53

4.6.3 Evaluasi Keterbatasan dan Persoalan Data	53
BAB V	55
PENUTUP.....	55
5.1 Kesimpulan	55
5.2 Saran	55
REFERENSI	57
LAMPIRAN.....	60



DAFTAR TABEL

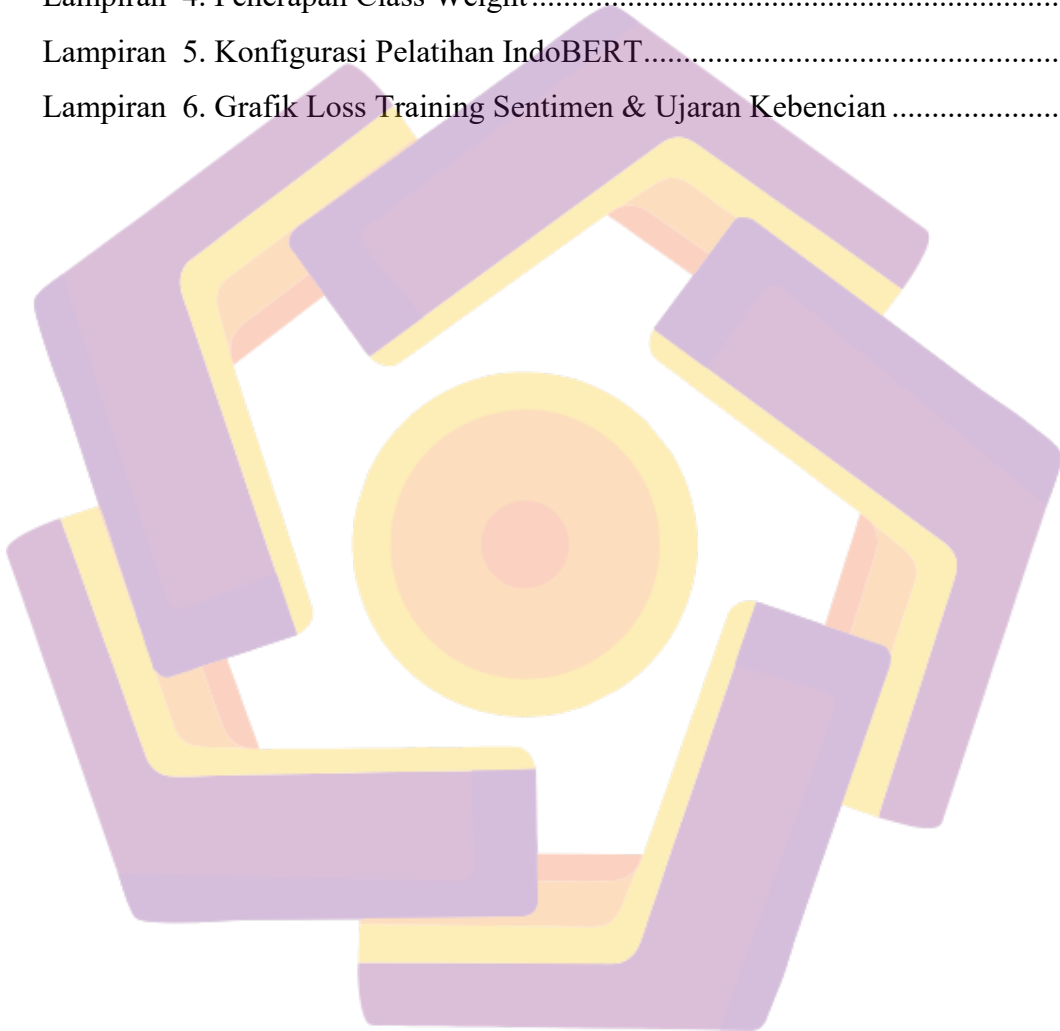
Tabel 3. 1 Contoh Hasil Pelabelan.....	22
Tabel 3. 2 Contoh Prapemrosesan	23
Tabel 3. 3 Perangkat Keras	28
Tabel 3. 4 Perangkat Lunak	29
Tabel 4. 1 Pembagian Data untuk Sentimen dan Deteksi Ujaran Kebencian	34
Tabel 4. 2 Hasil Evaluasi Majority Class Baseline.....	34
Tabel 4. 3 Hasil TF-IDF + Logistic Regression dengan Class Weighting	35
Tabel 4. 4 Hasil Evaluasi Majority Class Baseline.....	36
Tabel 4. 5 Hasil TF-IDF + Logistic Regression	36
Tabel 4. 6 Konfigurasi Pelatihan Model IndoBERT Analisis Sentimen	37
Tabel 4. 7 Bobot Kelas pada Sentimen.....	38
Tabel 4. 8 Hasil Evaluasi Model IndoBERT pada Data Uji	39
Tabel 4. 9 Contoh Kesalahan Klasifikasi Negatif diprediksi Netral.....	41
Tabel 4. 10 Contoh Kesalahan Klasifikasi Netral diprediksi Negatif.....	42
Tabel 4. 11 Contoh Kesalahan Klasifikasi Positif diprediksi Negatif	42
Tabel 4. 12 Konfigurasi Pelatihan Model IndoBERT Deteksi Ujaran Kebencian	43
Tabel 4. 13 Bobot Kelas pada Deteksi Ujaran Kebencian.....	44
Tabel 4. 14 Hasil Evaluasi Model IndoBERT	45
Tabel 4. 15 Contoh Kesalahan Klasifikasi False Positive	47
Tabel 4. 16 Contoh Kesalahan Klasifikasi False Negative.....	47
Tabel 4. 17 Distribusi Komentar Sentimen dan Ujaran Kebencian	49
Tabel 4. 18 Perbandingan Komentar Negatif dan Ujaran Kebencian.....	50
Tabel 4. 19 Perbandingan Model untuk Analisis Sentimen.....	52
Tabel 4. 20 Perbandingan Model untuk Deteksi Ujaran Kebencian.....	53

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	21
Gambar 4. 1 Distribusi Kelas dan Karakteristik Data.....	31
Gambar 4. 2 Distribusi Kelas Sentimen.....	31
Gambar 4. 3 Distribusi Kelas Ujaran Kebencian.....	32
Gambar 4. 4 Distribusi Panjang Komentar	33
Gambar 4. 5 Word Cloud.....	33
Gambar 4. 6 Hasil Pelatihan Model IndoBERT per Epoch	39
Gambar 4. 7 Hasil Confusion Matrix Sentimen.....	41
Gambar 4. 8 Hasil Model per Epoch.....	44
Gambar 4. 9 Confusion Matrix Deteksi Ujaran Kebencian	46
Gambar 4. 10 Distribusi Sentimen.....	49

DAFTAR LAMPIRAN

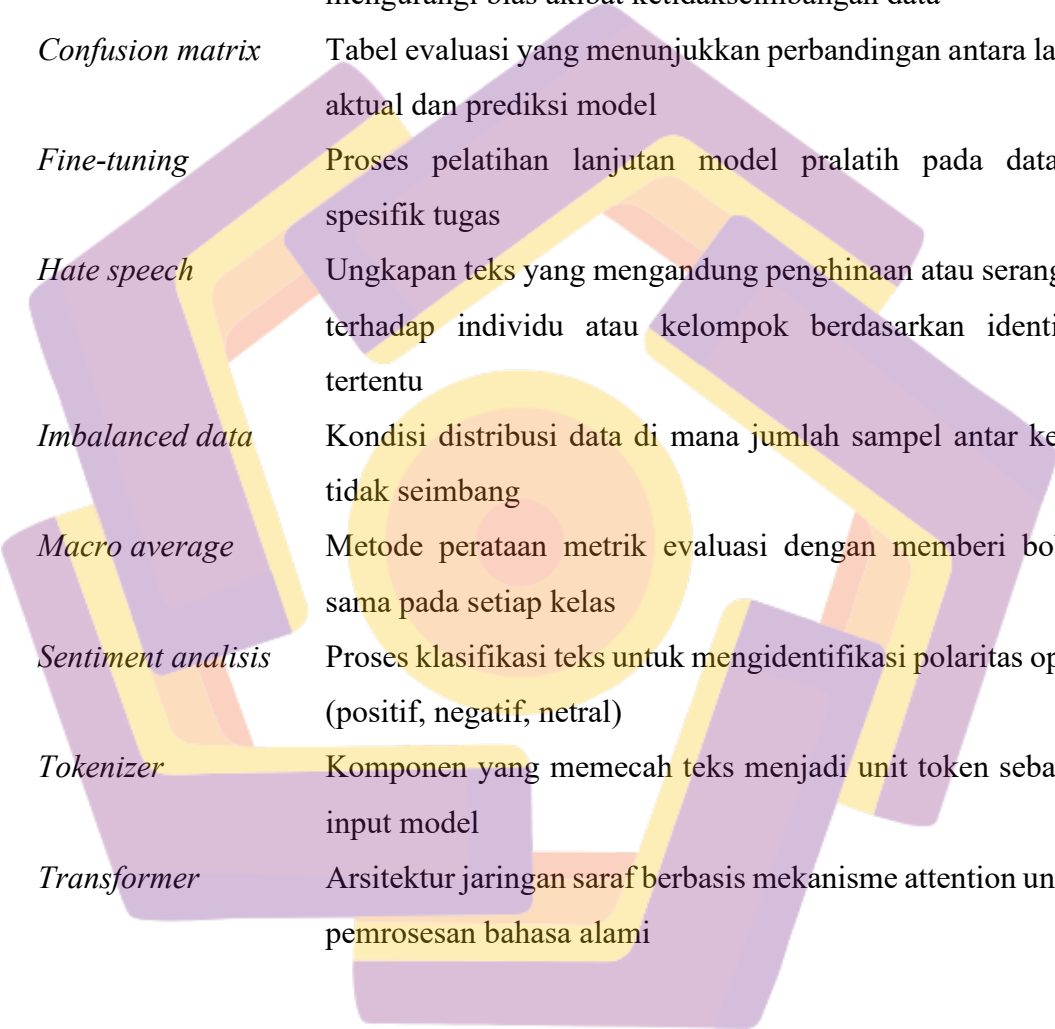
Lampiran 1. Data Komentar Hasil Scraping	60
Lampiran 2. Data Komentar Hasil Labeling	60
Lampiran 3. Data Komentar Setelah Preprocessing	61
Lampiran 4. Penerapan Class Weight	61
Lampiran 5. Konfigurasi Pelatihan IndoBERT	61
Lampiran 6. Grafik Loss Training Sentimen & Ujaran Kebencian	62



DAFTAR LAMBANG DAN SINGKATAN

API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CSV	Comma Separated Values
FP	False Positive
FN	False Negative
GPU	Graphics Processing Unit
IndoBERT	Indonesian Bidirectional Encoder Representations from Transformers
MLM	Masked Language Modeling
NLP	Natural Language Processing
TF-IDF	Term Frequency–Inverse Document Frequency
TP	True Positive
TN	True Negative

DAFTAR ISTILAH



<i>Baseline</i>	Model pembanding sederhana yang digunakan sebagai tolok ukur awal performa model utama
<i>Class weighting</i>	Teknik pemberian bobot berbeda pada tiap kelas untuk mengurangi bias akibat ketidakseimbangan data
<i>Confusion matrix</i>	Tabel evaluasi yang menunjukkan perbandingan antara label aktual dan prediksi model
<i>Fine-tuning</i>	Proses pelatihan lanjutan model pralatih pada dataset spesifik tugas
<i>Hate speech</i>	Ungkapan teks yang mengandung penghinaan atau serangan terhadap individu atau kelompok berdasarkan identitas tertentu
<i>Imbalanced data</i>	Kondisi distribusi data di mana jumlah sampel antar kelas tidak seimbang
<i>Macro average</i>	Metode perataan metrik evaluasi dengan memberi bobot sama pada setiap kelas
<i>Sentiment analysis</i>	Proses klasifikasi teks untuk mengidentifikasi polaritas opini (positif, negatif, netral)
<i>Tokenizer</i>	Komponen yang memecah teks menjadi unit token sebagai input model
<i>Transformer</i>	Arsitektur jaringan saraf berbasis mekanisme attention untuk pemrosesan bahasa alami

INTISARI

Kolom komentar YouTube pada topik sosial sensitif sering memuat beragam opini, mulai dari tanggapan kritis hingga ekspresi bernuansa kebencian, sehingga berpotensi menurunkan kualitas diskusi digital. Pada video Najwa Shihab “Susahnya Jadi Perempuan part 2”, tingginya volume komentar membuat klasifikasi manual tidak efisien dan rentan subjektivitas, sehingga diperlukan pendekatan analisis teks otomatis berbasis *Natural Language Processing*. Penelitian ini bertujuan menerapkan model IndoBERT untuk klasifikasi sentimen dan deteksi ujaran kebencian pada komentar berbahasa Indonesia pada video tersebut. Data dikumpulkan menggunakan YouTube Data API, diproses secara selektif, dilabeli untuk dua tugas klasifikasi, kemudian digunakan dalam proses *fine-tuning* IndoBERT dengan strategi *class-weighted loss* untuk menangani ketidakseimbangan kelas, serta dibandingkan dengan baseline Majority Class dan TF-IDF + Logistic Regression. Hasil pengujian menunjukkan IndoBERT melampaui baseline pada analisis sentimen dengan akurasi 80,0% dan F1-score makro 0,75. Pada deteksi ujaran kebencian, model mencapai akurasi 66,0% dengan F1-score kelas Hate 0,26 dan recall 0,52, yang menunjukkan kemampuan deteksi awal pada kelas minoritas yang tidak dicapai baseline. Distribusi hasil memperlihatkan sentimen negatif dominan (58,6%), namun hanya 14,3% di antaranya tergolong ujaran kebencian. Penelitian ini menjadi studi terapan penggunaan IndoBERT pada komentar YouTube bertopik isu gender dan dapat dijadikan rujukan metodologis untuk penelitian lanjutan pada domain serupa.

Kata kunci: analisis sentimen, ujaran kebencian, IndoBERT, YouTube, NLP

ABSTRACT

YouTube comment sections on sensitive social topics often contain diverse opinions, ranging from critical responses to hateful expressions, which may reduce the quality of digital discussions. In the video “Susahnya Jadi Perempuan part 2” by Najwa Shihab, the large volume of comments makes manual analysis inefficient and difficult to maintain consistently, creating the need for automated text analysis using Natural Language Processing. This study aims to apply the IndoBERT model for sentiment classification and hate speech detection on Indonesian-language YouTube comments from the selected video. Data were collected using the YouTube Data API, selectively preprocessed, and labeled for two classification tasks. The model was then fine-tuned using a class-weighted loss strategy to address class imbalance and compared with baseline models, namely Majority Class and TF-IDF with Logistic Regression. Experimental results show that IndoBERT outperforms the baselines in sentiment analysis with an accuracy of 80.0% and a macro F1-score of 0.75. For hate speech detection, the model achieves 66.0% accuracy, with a Hate-class F1-score of 0.26 and recall of 0.52, indicating initial detection capability for the minority class that baseline models fail to capture. The distribution results indicate that negative sentiment dominates (58.6%), while only 14.3% of those comments are categorized as hate speech. This work provides a case-study-based application of IndoBERT for sentiment and hate speech analysis on gender-related YouTube discussions and serves as a methodological reference for similar future studies.

Keyword: *sentiment analysis, hate speech detection, IndoBERT, YouTube comments, natural language processing*