



**ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP
JUDI ONLINE DI INDONESIA MENGGUNAKAN
FINE-TUNING INDOBERT**

SKRIPSI

disusun oleh :

ANASTASYA KAROLIEN GUNAWAN

22166013

**PROGRAM STUDI SARJANA SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS HARKAT NEGERI
2026**



SKRIPSI

**ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP
JUDI *ONLINE* DI INDONESIA MENGGUNAKAN
FINE-TUNING INDOBERT**

***SENTIMENT ANALYSIS OF TIKTOK COMMENTS ON ONLINE
GAMBLING IN INDONESIA USING FINE-TUNED INDOBERT***

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi

disusun oleh :

**ANASTASYA KAROLIEN GUNAWAN
22166013**

**PROGRAM STUDI SARJANA SISTEM INFORMASI
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS HARKAT NEGERI
2026**

HALAMAN PERNYATAAN KEASLIAN

Yang bertandatangan di bawah ini :

Nama : Anastasya Karolien Gunawan
NIM : 22166013
Jurusan / Program Studi : Sistem Informasi
Jenis Karya : Skripsi

Adalah mahasiswa Program Studi S1 Sistem Informasi, dengan ini saya menyatakan bahwa laporan Skripsi yang berjudul "**ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI *ONLINE* DI INDONESIA MENGGUNAKAN *FINE-TUNING* INDOBERT**" Merupakan hasil pemikiran dan kerjassama sendiri secara orisinil dan saya susun secara mandiri dan tidak melanggar kode etika hak karya cipta. Pada pelaporan Skripsi ini juga bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik tertentu di suatu perguruan tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau disebutkan dalam daftar pustaka.

Apabila di kemudian hari ternyata Laporan Skripsi ini terbukti melanggar kode etik karya cipta atau merupakan karya yang dikategorikan mengandung unsur plagiarisme, maka saya bersedia untuk melakukan penelitian baru dan menyusun laporannya sebagai Laporan Skripsi, sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan sesungguhnya.

Tegal, 3 Agustus 2026



(Anastasya Karolien Gunawan)

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai civitas akademika Universitas Harkat Negeri Tegal, Saya yang bertanda tangan dibawah ini :

Nama : Anastasya Karolien Gunawan
NIM : 22166013
Jurusan / Program Studi : Sistem Informasi
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Harkat Negeri Tegal **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas Skripsi saya yang berjudul :

ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI *ONLINE* DI
INDONESIA MENGGUNAKAN *FINE-TUNING* INDOBERT

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti *Noneksklusif* ini Universitas Harkat Negeri Tegal berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Kota Tegal

Pada Tanggal : 6 Juli 2026



(Anastasya Karolien Gunawan)

HALAMAN PERSETUJUAN

**ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI
ONLINE DI INDONESIA MENGGUNAKAN
FINE-TUNING INDOBERT**

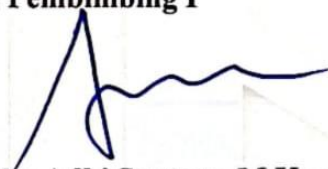
**SENTIMENT ANALYSIS OF TIKTOK COMMENTS ON ONLINE
GAMBLING IN INDONESIA USING FINE-TUNED INDOBERT**

yang disusun dan diajukan oleh

Anastasya Karolien Gunawan
22166013

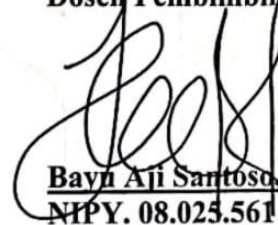
telah disetujui oleh Dosen Pembimbing SKRIPSI
pada tanggal 3 Agustus 2026

Dosen Pembimbing I



Nugroho Adhi Santoso, M.Kom
NIPY. 08.025.560

Dosen Pembimbing II



Bayu Aji Santoso, M.Kom
NIPY. 08.025.561

HALAMAN PENGESAHAN

**ANALISIS SENTIMEN KOMENTAR TIKTOK TERHADAP JUDI
ONLINE DI INDONESIA MENGGUNAKAN
FINE-TUNING INDOBERT**

**SENTIMENT ANALYSIS OF TIKTOK COMMENTS ON ONLINE
GAMBLING IN INDONESIA USING FINE-TUNED INDOBERT**

yang disusun dan diajukan oleh

Anastasya Karolien Gunawan

22166013

Telah dipertahankan di depan Dewan Penguji
pada tanggal 5 Agustus 2026

Susunan Dewan Penguji

Nama Penguji

**Sarif Surejo, S.E., M.Kom.
NIPY. 08.025.559**

**R. Bangkit Indarmawan Nugroho, M.Kom.
NIPY. 08.025.558**

**Bayu Aji Santoso, M.Kom.
NIPY. 08.025.561**

Tanda Tangan

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 5 Agustus 2026

Ketua Program Studi S1 Sistem Informasi

 **Universitas
Himpunan Sarjana
S1 Sistem Informasi**

**R. Bangkit Indarmawan N., M.Kom
NIPY. 08.025.558**

HALAMAN MOTO

Jangan menunggu sampai sempurna untuk memulai. Mulailah, dan biarkan proses yang menyempurnakanmu.

HALAMAN PERSEMBAHAN

Puji syukur ke hadirat Allah SWT atas segala rahmat, karunia, serta pertolongan-Nya, sehingga penulis dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk memperoleh gelar Sarjana. Setiap proses yang dilalui tidak terlepas dari doa dan dukungan serta kasih sayang yang selalu hadir di dalam proses dan langkah penulis. Oleh karena itu, dengan penuh rasa syukur dan hormat, karya sederhana ini penulis persembahkan kepada:

- 1 Papah tercinta, J.Romiko Gunawan, yang selalu memberikan doa dan dukungan yang tiada henti serta kasih sayang yang tak pernah putus.
- 2 Mamah tercinta, Mimin, yang selalu ada di samping penulis, memberikan dukungan serta doa yang tulus, serta menjadi penyemangat bagi penulis untuk terus berjuang hingga mampu menyelesaikan pendidikan ini.
- 3 Kedua adik tercinta, Evelina dan Vieri, terima kasih atas doa, perhatian, dan dukungan yang selalu hadir di setiap langkah penulis.
- 4 Kekasihku, Dwi Adit Prastyo, Terima kasih telah menjadi support system terbaik, pendengar yang sabar, dan alasan bagiku untuk tidak menyerah hingga skripsi ini selesai.
- 5 Keluarga Besar dan Saudara-saudaraku, Terima kasih atas dukungan, motivasi, dan doa yang selalu mengiringi setiap langkah penulis.
- 6 Bapak Nugroho Adhi Santoso, M.Kom dan Bapak Bayu Aji Santoso, M.Kom, Sebagai dosen pembimbing yang telah meluangkan waktu, memberikan arahan, dan membagikan ilmunya dengan sabar selama proses penyusunan skripsi ini.
- 7 Sahabat-sahabat, Terima kasih untuk setiap diskusi, tawa, dan dukungan moral di masa-masa sulit. Kita berhasil melewati ini bersama.
- 8 Dan untuk diriku sendiri, Terima kasih karena tidak pernah menyerah, tetap bertahan di saat lelah, dan percaya bahwa proses ini akan membawamu pada versi terbaik dari dirimu.

Tegal, 3 Agustus 2026

Anastasya Karolien Gunawan

INTISARI

Penelitian ini bertujuan untuk merancang pipeline klasifikasi sentimen, mengevaluasi kinerja model Transformer, dan memetakan pola opini publik mengenai isu judi online di Indonesia berdasarkan komentar TikTok. *Dataset* awal sebanyak 99.065 komentar disaring melalui tahap *preprocessing* (*cleaning*, normalisasi *slang*, serta filtering komentar pendek dan *bot-like*) hingga menghasilkan 73.782 data valid. Pelabelan otomatis dilakukan menggunakan pendekatan *Domain-Specific Lexicon* (*weak supervision*) yang mengategorikan data ke dalam kelas Netral (71,8%), Positif (19,9%), dan Negatif (8,3%). *Dataset* dibagi menggunakan *Stratified Sampling* dengan rasio 70:15:15. Untuk mengatasi masalah ketidakseimbangan kelas ekstrem (*extreme class imbalance*), proses *fine-tuning* model IndoBERT-base-p1 dioptimalkan menggunakan *Custom WeightedTrainer* dengan pembobotan kelas *inverse frequency*. Hasil evaluasi pada *test set* menunjukkan model IndoBERT mencapai *Accuracy* sebesar 99,28% dan *Macro F1-Score* sebesar 98,90%, secara signifikan mengungguli model *baseline* TF-IDF + *LinearSVC* (*Macro F1-Score* 93,72%). Capaian tinggi ini mencerminkan efektivitas model dalam mempelajari dan mereplikasi pola pelabelan leksikon (*weak supervision*), bukan keakuratan sentimen terhadap *ground truth* anotasi manusia asli. Pembahasan pola narasi mengindikasikan bahwa dominasi kelas Netral berpotensi merepresentasikan fenomena *information fatigue*, diikuti kelas Positif yang didominasi desakan pemblokiran infrastruktur digital, serta kelas Negatif yang menyoroti dampak sosial korban dan kritik penegakan hukum. Kesimpulannya, integrasi leksikon domain-spesifik dan *WeightedTrainer* pada IndoBERT terbukti *robust* dalam menangani teks informal yang tidak seimbang, serta memberikan referensi berbasis bukti bagi regulator dalam merumuskan kebijakan publik.

Kata kunci: Analisis Sentimen, IndoBERT, *Fine-Tuning*, Judi Online, TikTok.

ABSTRACT

This study aims to design a sentiment classification pipeline, evaluate Transformer model performance, and map public opinion patterns regarding online gambling in Indonesia based on TikTok comments. The initial dataset of 99,065 comments was filtered through preprocessing stages (cleaning, slang normalization, and filtering short and bot-like comments), resulting in 73,782 valid entries. Automatic proxy labeling was performed using a Domain-Specific Lexicon (weak supervision), categorizing the data into Neutral (71.8%), Positive (19.9%), and Negative (8.3%) classes. The dataset was split using Stratified Sampling with a 70:15:15 ratio. To handle extreme class imbalance, fine-tuning of the IndoBERT-base-p1 model was optimized using a Custom WeightedTrainer with inverse frequency class weighting. Evaluation results on the test set show that the IndoBERT model achieved 99.28% Accuracy and a 98.90% Macro F1-Score, significantly outperforming the TF-IDF + LinearSVC baseline model (Macro F1-Score 93.72%). This high performance reflects the model's effectiveness in learning and replicating lexicon-based labeling patterns rather than sentiment accuracy against human ground-truth annotations. Narrative analysis indicates that the dominance of the Neutral class potentially represents information fatigue, followed by the Positive class demanding digital infrastructure blocking, and the Negative class highlighting social impacts on victims and criticisms of law enforcement. In conclusion, integrating a domain-specific lexicon and WeightedTrainer in IndoBERT proves robust in handling imbalanced informal text, providing evidence-based insights for regulators in policy formulation.

Keyword: Sentiment Analysis, IndoBERT, Fine-Tuning, Online Gambling, TikTok.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas rahmat, hikmat, dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul **“Analisis Sentimen Komentar TikTok Terhadap Judi *Online* Di Indonesia Menggunakan Fine-Tuning IndoBERT”** tepat pada waktunya. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi di Universitas Harkat Negeri.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa tidak sedikit pihak yang telah memberikan bimbingan, dukungan, dan bantuan, baik secara moril maupun materil. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Sudirman Said, Ak., M.B.A selaku Rektor Universitas Harkat Negeri.
2. Bapak Dr. Irendra Radjawali, selaku Dekan Fakultas Sains dan Teknologi.
3. Bapak R. Bangkit Indarmawan N., M.Kom, selaku Ketua Program Studi Sarjana Sistem Informasi.
4. Bapak Nugroho Adhi Santoso, M.Kom, selaku Dosen Pembimbing 1 dan Bapak Bayu Aji Santoso, M.Kom, selaku Dosen Pembimbing 2 yang telah meluangkan waktu, pikiran, dan tenaga untuk memberikan bimbingan serta arahan selama proses penyusunan skripsi ini.
5. Bapak Dosen Penguji 1 dan Dosen Penguji 2 yang telah memberikan masukan dan saran yang membangun.
6. Kedua orang tua dan seluruh keluarga tercinta, yang selalu memberikan doa, motivasi, dan dukungan tanpa henti.
7. Teman-teman seperjuangan Program Studi Sistem Informasi angkatan 2022, serta semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu kelancaran penyusunan skripsi ini.
8. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah membantu kelancaran penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, baik dari segi penyusunan, bahasa, maupun kedalaman materi. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan skripsi ini.

Akhir kata, semoga skripsi ini dapat memberikan manfaat bagi pembaca, para akademisi, serta dapat menjadi referensi yang berguna bagi penelitian selanjutnya di bidang *Natural Language Processing* dan Sistem Informasi.

Tegal, 3 Agustus 2026

Anastasya Karolien G

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KEASLIAN	ii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
HALAMAN PENGESAHAN	v
HALAMAN MOTO	vi
HALAMAN PERSEMBAHAN	vii
INTISARI.....	viii
ABSTRACT.....	ix
KATA PENGANTAR	x
DAFTAR ISI.....	xii
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN.....	xvi
DAFTAR LAMBANG DAN SINGKATAN	xvii
DAFTAR ISTILAH.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	9
2.1 Landasan Teori.....	9
2.2 Penelitian Terdahulu	23
BAB III METODE PENELITIAN	26
3.1 Pengumpulan Data	26
3.2 <i>Preprocessing</i> dan Pelabelan Data.....	31
3.3 Pembagian Data <i>Training</i> dan <i>Testing</i>	37
3.4 Model Pembandingan (Baseline: TF-IDF + <i>LinearSVC</i>).....	41

3.5 Pelatihan Model	43
3.6 Metrik Evaluasi	48
3.7 Alur Penelitian	50
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	53
4.1 Deskripsi <i>Dataset</i>	53
4.2 <i>Preprocessing</i> dan Pelabelan Data.....	55
4.3 Pembagian Data <i>Training</i> dan <i>Testing</i>	61
4.4 Hasil Eksperimen Model.....	62
4.5 Analisis Pola Narasi dan Diskursus Publik.....	64
4.6 Pembahasan.....	67
BAB V PENUTUP	71
DAFTAR PUSTAKA	74
LAMPIRAN.....	79

DAFTAR TABEL

Tabel 2. 1 Karakteristik Data Komentar TikTok untuk Analisis Sentimen	10
Tabel 2. 2 Pipeline Preprocessing Teks Komentar TikTok	13
Tabel 2. 3 Perbandingan BERT vs. Word Embedding Statis	14
Tabel 2. 4 Perbandingan Performa Model Sentimen Bahasa Indonesia	18
Tabel 2. 5 Hyperparameter Rekomendasi untuk Fine-Tuning IndoBERT	19
Tabel 2. 6 Contoh Confusion Matrix 3-Kelas	21
Tabel 2. 7 Interpretasi Nilai Metrik Evaluasi	23
Tabel 2. 8 Matriks literatur review dan posisi penelitian	24
Tabel 3. 1 Dataset Awal	27
Tabel 3. 2 Spesifikasi dan Karakteristik Dataset Penelitian	29
Tabel 3. 3 Tahapan Preprocessing dan Implementasi Teknis	32
Tabel 3. 4 Pelabelan Sentimen Menggunakan Pendekatan Lexicon-Based	36
Tabel 3. 5 Distribusi Kelas Sentimen Hasil Pelabelan Leksikon	37
Tabel 3. 6 Pembagian Dataset untuk Pelatihan dan Evaluasi Model	39
Tabel 3. 7 Bobot Kelas (Class Weights) untuk Custom WeightedTrainer	45
Tabel 3. 8 Konfigurasi Hiperparameter Pelatihan IndoBERT	46
Tabel 3. 9 Rumus dan Interpretasi Metrik Evaluasi	50
Tabel 4.1 Meta <i>Dataset</i> Mentah	53
Tabel 4. 2 Ringkasan Reduksi Data Preprocessing	55
Tabel 4. 3 Hasil dari Preprocessing Data	56
Tabel 4.4 Hasil Proses Scoring Secara Granular	59
Tabel 4.5 Pembagian <i>Dataset</i> untuk Pelatihan dan Evaluasi Model	62
Tabel 4.6 Perbandingan Performa Model Baseline dan IndoBERT pada Test Set63	

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur BERT untuk Klasifikasi Teks	15
Gambar 2. 2 Diagram Alur <i>Fine-Tuning</i> IndoBERT.....	21
Gambar 3. 1 Alur Penelitian	50
Gambar 4. 1 Distribusi Sentimen Publik TikTok - Isu Judi Online.....	60
Gambar 4. 2 <i>Confusion Matrix</i> Model IndoBERT pada <i>Test Set</i>	64
Gambar 4. 3 <i>Word Cloud</i> per Kelas Sentimen.....	65

DAFTAR LAMPIRAN

Lampiran A- 1 Dataset Awal judul_tiktok.csv 1	79
Lampiran B- 1 Contoh Dataset Setelah <i>Preprocessing</i>	80
Lampiran C- 1 Hasil Perhitungan Leksikon Domain-Spesifik	81
Lampiran D- 1 Statistik & Distribusi Dataset	82
Lampiran E- 1 Dokumentasi <i>Source Code (Script Python)</i>	83
Lampiran F- 1 Dokumentasi Lingkungan Komputasi	90
Lampiran G- 1 Dokumentasi <i>Source Code (Script Python)</i>	91
Lampiran H- 1 <i>Link Script Google Colab</i>	96
Lampiran I- 1 Jadwal Kegiatan	97
Lampiran J- 1 Lembar Bimbingan	98
Lampiran K- 1 Letter of Acceptance (LoA)	102
Lampiran L- 1 Formulir Perubahan Judul	103
Lampiran M- 1 Bukti Cek Plagiarisme	104
Lampiran N- 1 <i>Curriculum Vitae</i>	106
Lampiran O- 1 Formulir Publikasi Jurnal	107

DAFTAR LAMBANG DAN SINGKATAN

Lambang / Singkatan	Kepanjangan / Keterangan
$h_{[CLS]}$	Vektor representasi token [CLS] (berdimensi 768)
W	Matriks bobot pada lapisan klasifikasi linier
b	Vektor bias pada lapisan klasifikasi linier
N	Total jumlah data pada <i>training set</i>
k	Jumlah kelas sentimen (bernilai 3)
n_i	Jumlah data pada kelas sentimen ke- i
$\mathcal{L}$	Fungsi kerugian (<i>Cross-Entropy Loss</i>)
y_i	Label aktual (<i>ground truth</i>) dari data ke- i
$\hat{y}_i$	Probabilitas prediksi model untuk data ke- i
LLM	<i>Large Language Model</i>
MLM	<i>Masked Language Modeling</i>
NLP	<i>Natural Language Processing</i>
NSP	<i>Next Sentence Prediction</i>
OJK	Otoritas Jasa Keuangan
Polri	Kepolisian Negara Republik Indonesia
PPATK	Pusat Pelaporan dan Analisis Transaksi Keuangan
SOTA	<i>State-of-the-Art</i>
SVM	<i>Support Vector Machine</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
URL	<i>Uniform Resource Locator</i>

DAFTAR ISTILAH

Analisis Sentimen	Teknik komputasi dalam Natural Language Processing (NLP) untuk mengidentifikasi, mengekstraksi, dan mengkategorikan opini atau emosi dalam teks ke dalam kelas polaritas tertentu (positif, negatif, netral).
<i>Class Imbalance</i>	Kondisi ketidakseimbangan distribusi data antar kelas dalam <i>dataset</i> , di mana satu kelas mendominasi secara jumlah dibandingkan kelas lainnya, yang dapat menyebabkan model menjadi bias.
<i>Code-mixing</i>	Fenomena pencampuran dua atau lebih bahasa dalam satu tuturan atau kalimat, sangat umum ditemukan pada teks media sosial Indonesia (misalnya campuran Bahasa Indonesia, Jawa, dan Inggris).
<i>Confusion Matrix</i>	Tabel evaluasi performa model klasifikasi yang membandingkan hasil prediksi model dengan label aktual (ground truth) untuk menghitung metrik seperti <i>precision</i> , <i>recall</i> , dan <i>F1-Score</i> .
<i>Deep Learning</i>	Sub-bidang dari machine learning yang menggunakan jaringan saraf tiruan berlapis-lapis (<i>neural networks</i>) untuk memproses data dan mengenali pola kompleks secara otomatis.
<i>Design Science Research (DSR)</i>	Paradigma penelitian yang berfokus pada pengembangan, implementasi, dan evaluasi artefak komputasi (seperti model atau sistem) sebagai solusi untuk memecahkan masalah nyata.
<i>Domain-Specific Lexicon</i>	Kamus kata yang dikonstruksi secara khusus untuk domain atau topik tertentu (dalam hal ini, judi online), bukan kamus sentimen umum, agar mampu menangkap kosakata spesifik dan konteks lokal.

<i>F1-Score</i>	Metrik evaluasi yang merupakan rata-rata harmonik dari precision dan recall, digunakan untuk mengukur akurasi model secara seimbang, terutama pada <i>dataset</i> yang tidak seimbang (<i>imbalanced</i>).
<i>Fine-Tuning</i>	Proses melatih kembali model pra-latih (<i>pre-trained model</i>) pada <i>dataset</i> spesifik dengan laju pembelajaran (<i>learning rate</i>) kecil untuk mengadaptasi representasi semantik model ke tugas baru tanpa menghilangkan pengetahuan sebelumnya.
IndoBERT	Model bahasa pra-latih berbasis arsitektur Transformer yang dikembangkan khusus untuk memahami bahasa Indonesia, dilatih pada korpus teks Indonesia yang masif.
Judi Online (Judol)	Aktivitas perjudian yang dilakukan melalui platform digital atau internet, menjadi objek kajian utama dan domain spesifik dalam penelitian ini.
<i>Machine Learning</i>	Cabang kecerdasan buatan yang memungkinkan sistem komputer belajar dan meningkatkan performanya dari data secara otomatis tanpa diprogram secara eksplisit.
<i>Macro F1-Score</i>	Rata-rata <i>F1-Score</i> yang dihitung dengan memberikan bobot yang sama untuk setiap kelas, tanpa memperhitungkan ukuran atau jumlah data masing-masing kelas, sehingga adil untuk data yang tidak seimbang.
<i>Natural Language Processing (NLP)</i>	Bidang ilmu komputer dan linguistik komputasi yang berfokus pada interaksi antara komputer dan bahasa manusia, termasuk pemahaman, analisis, dan generasi teks.
<i>Preprocessing</i>	Tahap awal pengolahan data mentah menjadi format yang bersih, terstruktur, dan konsisten agar siap diproses oleh model <i>machine learning</i> .
<i>Proxy Label</i>	Label data yang dihasilkan secara otomatis melalui aturan atau heuristik (seperti leksikon) sebagai pengganti label

manual, digunakan dalam pendekatan *weak supervision* untuk *dataset* berskala besar.

<i>Self-Attention</i>	Mekanisme inti dalam arsitektur Transformer yang memungkinkan model menimbang pentingnya setiap kata dalam kalimat terhadap kata lainnya untuk menangkap konteks bidireksional dan dependensi jarak jauh.
Slang	Kata, frasa, atau ungkapan tidak baku yang digunakan dalam percakapan informal, sangat dominan dan dinamis pada teks media sosial seperti TikTok.
<i>State-of-the-Art (SOTA)</i>	Tingkat pencapaian tertinggi atau metode paling mutakhir dan terbaik yang ada pada suatu bidang penelitian pada saat tertentu.
<i>Stratified Sampling</i>	Teknik pembagian <i>dataset</i> yang mempertahankan persentase proporsi kelas yang sama pada setiap subset (<i>training</i> , <i>validation</i> , <i>test</i>) untuk mencegah bias distribusi selama pelatihan.
<i>Tokenization</i>	Proses memecah teks menjadi unit-unit kecil yang disebut token (bisa berupa kata, sub-kata, atau karakter) sesuai dengan vokabular model agar dapat diproses secara numerik.
Transformer	Arsitektur jaringan saraf tiruan modern yang mengandalkan mekanisme attention untuk memproses data sekuensial seperti teks, menggantikan arsitektur RNN/LSTM tradisional.
<i>Weak Supervision</i>	Pendekatan pelatihan model menggunakan label yang dibuat secara otomatis, tidak sempurna, atau bersifat proksi (seperti dari leksikon) karena keterbatasan ketersediaan data berlabel manual yang akurat.
<i>WeightedTrainer</i>	Modifikasi kustom pada kelas Trainer di pustaka Hugging Face yang menyuntikkan bobot kelas (<i>class weights</i>) ke dalam fungsi kerugian (<i>loss function</i>) untuk menangani masalah class imbalance secara efektif.