

**LEMBAR REVISI SIDANG PROPOSAL SKRIPSI**

Semester Genap 2024/2025


**Nama** : MARVEL WILBERT ODELIO  
**NRP** : C14220223  
**Judul Skripsi** : Penerapan Model IndoBERTweet untuk Deteksi Komentar Promosi Judi Online Berbahasa Indonesia di Sosial Media  
**Tanggal Ujian** : 5 Juni 2025 Pukul : 11:45  
**Penguji** : 03024 - Liliana, S.T., M.Eng., Ph.D.

No	Materi yang di Revisi
1	cari justifikasi pakar untuk pelabelan dataset. tuliskan di ruang lingkup dengan disebutkan nama pakarnya.
2	belum terlihat research gap nya. salah satu saran saya adalah penambahan data yang mengandung kata-kata promosi yang belum ada di dataset sebelumnya. tuliskan dengan detail mekanisme pengumpulan data dan pelabelannya. pastikan ada pakar yang kredibel untuk memverifikasi label data dataset yang dikumpulkan tersebut → cari referensi dari penelitian linguistik.
3	bandingkan beberapa model yang apple to apple, dari beberapa kategori - pertimbangkan apakah

Surabaya, 10 Juni 2025  
Dosen Penguji,

CNN - fast text cocok?  
 - GNN, pastikan lagi  
 - TF-IDF spt-nya kurang  
 cocok untuk text pendek.

Liliana, S.T., M.Eng., Ph.D.  
 Nip : 03024

Kontrol Oleh	Keterangan	Paraf
Pembimbing I dan II	<u>Sudah</u> / Belum Perbaikan	

**LEMBAR REVISI SIDANG PROPOSAL SKRIPSI**


Semester Genap 2024/2025

**Nama** : MARVEL WILBERT ODELIO  
**NRP** : C14220223  
**Judul Skripsi** : Penerapan Model IndoBERTweet untuk Deteksi Komentar Promosi Judi Online Berbahasa Indonesia di Sosial Media  
**Tanggal Ujian** : 5 Juni 2025 Pukul : 11:45  
**Penguji** : 02030 - Dr. Gregorius Satia Budhi, S.T., M.T.

No	Materi yang di Revisi
	Data tidak ditemukan

Surabaya, 13 Juni 2025  
Dosen Penguji,

Dr. Gregorius Satia Budhi, S.T., M.T.  
Nip : 02030

Kontrol Oleh	Keterangan	Paraf
Pembimbing I dan II	Sudah / Belum Perbaikan	

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KRISTEN PETRA**

---

**USULAN PROYEK AKHIR**

Nama : Marvel Wilbert Odelio  
NRP : C14220223  
Program : Informatika  
Judul Skripsi : Penerapan Model IndoBERTweet untuk Deteksi Komentar Promosi Judi Online Berbahasa Indonesia di Sosial Media  
Pembimbing 1 : Hans Juwiantho, S.Kom., M.Kom.  
Pembimbing 2 : -  
Dilaksanakan : Gasal 2025 / 2026

13 - 06 - 2025

Yang mengusulkan



Marvel Wilbert Odelio

Mengetahui:

Pembimbing 1



Hans Juwiantho, S.Kom., M.Kom.

Koordinator Pra Skripsi



Liliana, S.T., M.Eng., Ph.D.

Transkrip

Nama

:

MARVEL WILBERT ODELIO

NRP

:

C14220223

Tempat/Tgl. Lahir

:

SURABAYA, 24/12/2003

Tahun Masuk

:

1-2022/2023

Fakultas

:

TEKNOLOGI INDUSTRI

Program Studi

:

INFORMATIKA

Program

:

INFORMATIKA

Program Pendidikan

:

S1

Kode	Mata Kuliah	SMT	SKS	Nilai
DU4197	AGAMA DAN HIDUP BERMAKNA / A MEANINGFUL LIFE	1-22/23	4	A
DU4122	BAHASA INDONESIA / INDONESIAN	1-22/23	2	A
TF4537	KONSEP ALGORITMA / ALGORITHM CONCEPTS	1-22/23	3	B+
TF4205	DASAR SISTEM KOMPUTER / INTRODUCTION TO COMPUTER SYSTEM	1-22/23	2	A
TF4536	DASAR PEMROGRAMAN / PROGRAMMING BASICS	1-22/23	4	A
FD4505	KALKULUS I / CALCULUS I	1-22/23	3	A
TF4220	PENGANTAR MANAJEMEN DAN BISNIS / INTRODUCTION TO MANAGEMENT AND BUSINESS	1-22/23	2	A
DU4101	PANCASILA / STATE IDEOLOGY	2-22/23	2	A
TF4227	STATISTIKA DASAR / BASIC STATISTICS	2-22/23	3	A
TF4267	KOMUNIKASI INTERPERSONAL / INTERPERSONAL COMMUNICATION	2-22/23	3	B+
TF4245	MATEMATIKA DISKRIT / DISCRETE MATHEMATICS	2-22/23	3	A
TF4235	PEMROGRAMAN BERORIENTASI OBYEK / OBJECT-ORIENTED PROGRAMMING	2-22/23	3	A
TF4253	JARINGAN KOMPUTER / COMPUTER NETWORKS	2-22/23	3	A
TF4229	BASIS DATA / DATABASE	2-22/23	3	B+
TF4504	BAHASA INGGRIS / ENGLISH	1-23/24	2	A
TF4249	PENGANTAR AKUNTANSI / INTRODUCTION TO ACCOUNTING	1-23/24	2	A
TF4255	REKAYASA PERANGKAT LUNAK / SOFTWARE ENGINEERING	1-23/24	3	B+
FD4507	ALJABAR LINIER / LINEAR ALGEBRA	1-23/24	3	A
TF4270	DESAIN DAN ANALISIS ALGORITMA / DESIGN AND ANALYSIS OF ALGORITHMS	1-23/24	3	A
TF4372	ARSITEKTUR DAN ORGANISASI KOMPUTER / COMPUTER ARCHITECTURES AND ORGANIZATION	1-23/24	3	A

Kode	Mata Kuliah	SMT	SKS	Nilai
DU4164	PENDIDIKAN KEWARGANEGARAAN / CIVICS EDUCATION	1-23/24	2	A
TF4343	TEKNOLOGI WEB / WEB TECHNOLOGIES	1-23/24	3	A
TF4219	STRUKTUR DATA / BASIC STUCTURES	1-23/24	3	A
DU4198	DIGITAL LEADERSHIP / DIGITAL LEADERSHIP	2-23/24	2	A
DU4163	ETIKA PROFESI / PROFESSIONAL ETHICS	2-23/24	2	A
TF4327	ANALISIS DAN DESAIN SISTEM INFORMASI / ANALYSIS AND DESIGN OF INFORMATION SYSTEM	2-23/24	3	B+
TF4409	KECERDASAN BUATAN / ARTIFICIAL INTELLIGENCE	2-23/24	3	A
TF4544	CYBER OPERATIONS / CYBER OPERATIONS	2-23/24	3	A
TF4415	GRAFIKA KOMPUTER / COMPUTER GRAPHICS	2-23/24	3	A
TF4243	SISTEM OPERASI / OPERATING SYSTEM	2-23/24	3	A
TF4247	METODE NUMERIK / NUMERICAL METHODS	2-23/24	2	A
TF4237	INTERAKSI MANUSIA DAN KOMPUTER / HUMAN-COMPUTER INTERACTION	2-23/24	3	A
FD4508	TECHNOPRENEURSHIP / TECHNOPRENEURSHIP	1-24/25	3	A
TF4507	SISTEM TERDISTRIBUSI / DISTRIBUTED SYSTEMS	1-24/25	3	B+
TF4275	COMPUTER VISION / COMPUTER VISION	1-24/25	3	A
TF4317	MANAJEMEN PROYEK TEKNOLOGI INFORMASI / INFORMATION TECHNOLOGY PROJECT MANAGEMENT	1-24/25	3	A
TF4581	GENERATIVE ARTIFICIAL INTELLIGENCE / GENERATIVE ARTIFICIAL INTELLIGENCE	1-24/25	3	A
TF4569	NETWORK DEFENSE / NETWORK DEFENSE	1-24/25	3	B+
TF4579	WEB FRAMEWORKS AND DEPLOYMENT / WEB FRAMEWORKS AND DEPLOYMENT	1-24/25	3	A

Keterangan :	
Indeks Prestasi Kumulatif	: 3.9
Jumlah SKS	: 109 SKS

Nilai Huruf	Keterangan	Nilai Bobot
A	Istimewa	4.00
B+	Baik Sekali	3.50
B	Baik	3.00
C+	Cukup Baik	2.50
C	Cukup	2.00
D	Kurang	1.00
E	Sangat Kurang	0.00

**Penerapan Model IndoBERTWeet untuk Deteksi Komentar Promosi Judi  
Online Berbahasa Indonesia di Sosial Media.**

**Proposal Skripsi**

Oleh:  
Marvel Wilbert Odelio  
NRP: c14220223

PROGRAM STUDI INFORMATIKA



**FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS KRISTEN PETRA  
SURABAYA  
2025**

## 1. PENDAHULUAN

### 1.1. Latar Belakang

#### 1.1.1. Analisis Masalah

Media sosial kini menjadi sarana utama komunikasi di era digital. Di Indonesia, pengguna internet menghabiskan rata-rata 3 jam 11 menit per hari di media sosial, melebihi rata-rata global (Mufrida, 2024). Namun, tingginya penggunaan ini juga dimanfaatkan untuk menyebarkan promosi judi online yang kian kompleks dan tersembunyi. Konten promosi sering disamarkan dengan istilah positif seperti *cuan*, *untung*, atau *slot gacor*, serta disebar melalui komentar pendek dari akun palsu pada unggahan viral (Perdana et al., 2024).

Taktik yang terus berkembang membuat pendekatan manual tak lagi efektif. Karena itu, diperlukan teknologi berbasis kecerdasan buatan, seperti *machine learning* dan *deep learning*, untuk mendeteksi promosi judi online secara efisien dan dalam skala besar.

Urgensi penanganan meningkat seiring masifnya konten terkait. Kemkomdigi mencatat lebih dari 1,3 juta konten negatif ditindak antara Oktober 2024 hingga Maret 2025, termasuk lebih dari 1,1 juta terkait judi online (*Apresiasi Laporan Masyarakat, Komdigi Tangani 1,3 Juta Konten Pornografi Dan Judi Online*, 2025). Dalam satu pekan di Januari 2025, 43.063 konten judi online juga ditangani (Rauf & US, 2025). Dampaknya tidak hanya digital, tetapi juga sosial dan ekonomi menyebabkan kriminalitas, konflik keluarga, gangguan mental, dan melemahkan fondasi ekonomi nasional (Selpiyani, 2024; Setiadi, 2024).

#### 1.1.2. Related Research

Penelitian oleh Perdana et al. (2024) merupakan studi awal yang mencoba mendeteksi promosi judi online di Twitter menggunakan pendekatan text mining klasik. Kombinasi TF-IDF dan Random Forest menjadi metode terbaik dalam penelitian tersebut, dengan presisi 96.6% dan recall 95.8%. Namun, pendekatan ini hanya mengandalkan empat kata kunci eksplisit ('slot', 'gacor', 'cuan', 'untung') dalam pengumpulan data, sehingga berisiko melewatkan promosi implisit yang menggunakan kata-kata tersamar maupun kata kunci lainnya.

Sejauh ini, studi yang secara khusus menargetkan deteksi promosi judi online, terutama dalam konteks bahasa Indonesia dan media sosial, masih sangat terbatas. Oleh karena itu, penting untuk melihat pendekatan lain yang telah terbukti efektif dalam tugas klasifikasi teks serupa seperti deteksi ujaran kebencian, sarkasme, berita palsu, dan analisis sentimen yang umumnya juga melibatkan konteks implisit dan bahasa informal. Dalam hal ini,

pendekatan deep learning mulai digunakan secara luas karena kemampuannya mengenali pola semantik dalam teks. Salah satu contohnya adalah penggunaan *Convolutional Neural Network* (CNN), yang terbukti efektif dalam menangkap frasa penting dalam suatu kalimat. Penelitian oleh Handoko et al., (2025) menunjukkan CNN dengan *embedding* FastText mampu mencapai akurasi 95.98% dalam tugas analisis sentimen, namun cenderung buruk dalam memahami konteks sekuensial yang kompleks. Sebaliknya, *Long Short Term Memory* (LSTM) dan variannya *Bidirectional Long Short Term Memory* (BiLSTM) lebih unggul dalam memahami urutan kata karena memiliki memori jangka panjang. Namun, LSTM hanya dapat memproses teks dalam satu arah. Berbeda dengan *Bidirectional LSTM* (BiLSTM) yang dapat memproses teks dari dua arah sehingga mampu membangun representasi yang jauh lebih kaya untuk setiap kata berdasarkan konteks penuh. Keunggulan ini terbukti atas performanya dalam deteksi depresi dimana akurasinya mencapai 98.45% (Huda et al., 2024).

Selain pendekatan *deep learning*, model-model klasik seperti *support vector machine* (SVM) dan *random forest* juga menunjukkan performa yang kompetitif ketika dipadukan dengan representasi fitur yang lebih kaya, seperti *word embedding* (Nuha & Lin, 2025). Li & Li, (2025) mencatat peningkatan akurasi hingga 92.84% dalam deteksi ujaran kebencian implisit dengan SVM dan Word2Vec. Pendekatan hibrida serupa juga berhasil diterapkan oleh Sibarani et al., (2025), di mana *random forest* yang awalnya hanya mencapai akurasi 79% melonjak hingga 97.24% saat diberi fitur *embedding* dari BERT. Studi serupa oleh Khaire et al., (2025) dan Kaware (2025) menunjukkan efektivitas pendekatan hibrida yang menggabungkan *embedding* kontekstual dari BERT dengan model klasik seperti SVM dan *random forest*. Khaire menggunakan DistilBERT untuk deteksi berita palsu, sedangkan Kaware mengembangkan Indo-HateSpeech, yang mengklasifikasikan ujaran kebencian di Instagram ke dalam tiga tingkat (No Hate, Hate, Extreme Hate) menggunakan fitur BERT dan model klasik. Keduanya menunjukkan bahwa kombinasi ini meningkatkan akurasi, F1-score, dan recall secara signifikan dibanding pendekatan tradisional. Hal ini memperkuat fakta bahwa representasi kontekstual dari transformer bisa dikombinasikan dengan model klasik secara fleksibel untuk menangani tugas-tugas klasifikasi teks yang kompleks.

Dalam konteks bahasa Indonesia, model pralatih seperti IndoBERT telah digunakan dalam berbagai tugas klasifikasi, namun keterbatasannya terletak pada data pelatihan yang bersifat formal (Nabiilah et al., 2023). Untuk menangani bahasa informal di media sosial, IndoBERTweet dikembangkan oleh Koto et al. (2021) dengan pelatihan pada lebih dari 26

juta tweet, membuatnya unggul dalam memahami slang dan struktur kalimat khas Twitter. Kusuma & Chowanda (2023) menunjukkan bahwa IndoBERTweet dengan BiLSTM mampu mencapai akurasi hingga 93.7% dalam mendeteksi ujaran kebencian. Hal ini memperkuat potensi IndoBERTweet untuk digunakan dalam tugas yang memerlukan pemahaman konteks implisit dan informal, seperti deteksi promosi judi online.

Tabel state-of-the-art

Nama Peneliti	Tahun	Masalah	Metode	Hasil	Perbedaan
Perdana et al.	2024	Deteksi promosi judi online di <i>Twitter</i> berbahasa Indonesia	<i>Random Forest, Logistic Regression, CNN dengan Word2Vec, TF-IDF, GloVe</i>	<i>TF-IDF + Random Forest</i> memiliki performa terbaik dengan presisi mencapai 96.6% dan recall mencapai 95.8%	Fokus deteksi pada tweet yang mengandung promosi judi online. Dataset hanya berasal dari twitter dan dikumpulkan dengan 4 keyword pencarian
Hidayatullah et al.	2023	Identifikasi konten negative pada tweet Indonesia	MBERT, IndoBERT, XLM-RoBERTa, IndoBERTweet	IndoBERTweet memiliki performa terbaik dengan Presisi 97.03%, Recall 96.88% dan F1-score mencapai 96.94%	Fokus pada mengenali konten negative seperti konten kekerasan dan konten pornografi
Kusuma & Chowanda	2023	Deteksi ujaran kebencian di <i>Twitter</i>	IndoBERTweet + BiLSTM	IndoBERTweet + BiLSTM: semua metrik mencapai 93%	Fokusnya pada deteksi ujaran kebencian, bukan promosi judi online
Sibarani et al.	2025	Deteksi Sarkasme	<i>Random Forest</i> dikombinasikan dengan IndoBERT	Performa meningkat dari 79% ke 97.24% untuk semua metrik	Fokusnya pada deteksi sarkasme, bukan promosi judi online
Li & Li	2025	Deteksi Ujaran Kebencian pada berbagai platform sosial media	SVM, BERT, HateBERT, <i>Random Forest</i>	SVM memiliki performa terbaik dengan akurasi lebih dari 90% untuk 2 dataset yang berbeda	Fokusnya pada deteksi ujaran kebencian, bukan promosi judi online

1.1.3. Analisis Solusi



Berdasarkan permasalahan yang diangkat dan studi sebelumnya, penelitian ini mengusulkan pengujian dan perbandingan lima metode klasifikasi. Setiap metode dipilih untuk merepresentasikan pendekatan yang berbeda secara fundamental dalam NLP, mulai dari machine learning klasik, hibrida, *deep learning*, hingga model *transformer*. Tujuannya adalah untuk mengevaluasi efektivitas masing-masing pendekatan dalam mendeteksi promosi judi online yang cenderung menggunakan bahasa tidak baku, slang, dan konteks implisit. Pemilihan metode didasarkan pada evolusi kemampuan model dalam merepresentasikan teks, dari yang paling sederhana hingga yang paling kompleks secara semantik:

**a. Pendekatan Leksikal (Baseline): Random Forest + TF-IDF**

Pendekatan ini jadi titik awal (baseline) karena sudah terbukti efektif di penelitian sebelumnya. Cara kerjanya cukup sederhana—mengandalkan seberapa sering kata-kata tertentu (seperti “slot”, “gacor”, atau “bonus”) muncul dalam teks.

**b. Pendekatan Sekuensial: IndoBERTweet + BiLSTM**

Pendekatan ini dipilih karena mampu menggabungkan pemahaman konteks bahasa informal khas Twitter dari IndoBERTweet dengan kemampuan BiLSTM dalam menangkap pola urutan kata secara sekuensial. Selain itu, metode ini telah terbukti memiliki performa yang baik dalam tugas klasifikasi ujaran kebencian seperti ditunjukkan dalam studi oleh (Kusuma & Chowanda, 2023).

**c. Pendekatan Hibrida: Random Forest/SVM + IndoBERTweet**

Pada pendekatan ini model klasik seperti Random Forest dan SVM diberi “pemahaman” yang lebih dalam lewat fitur dari IndoBERTweet. Jadi, model *transformer* menangkap makna kontekstual, lalu model klasik yang menentukan apakah teks tersebut mengandung promosi judi.

**d. Pendekatan Kontekstual End-to-End: Fine-Tuned IndoBERTweet**

Pendekatan ini merupakan pendekatan paling mutakhir. Seluruh model IndoBERTweet dilatih ulang khusus untuk mendeteksi promosi judi online. Karena sudah terbiasa dengan gaya bahasa Twitter Indonesia, model ini paling siap menangani bahasa tidak baku, slang, dan konteks tersembunyi.

Berdasarkan karakteristik tiap pendekatan, hipotesis penelitian ini dirumuskan sebagai berikut:

- a. Hipotesis Utama (H1): Pendekatan yang menggunakan representasi semantik-kontekstual (berbasis Transformer) akan secara signifikan mengungguli pendekatan leksikal dan pola lokal dalam klasifikasi promosi judi online.

Prediksi Urutan Performa (berdasarkan F1 Score):

- Fine-Tuned IndoBERTweet
- RF/SVM + IndoBERTweet
- IndoBERTweet + BiLSTM
- Random Forest + TF-IDF

- b. Hipotesis Nol (H0): Tidak terdapat perbedaan signifikan antara performa pendekatan kontekstual dan non-kontekstual.

Penelitian ini akan menguji hipotesis di atas dengan membandingkan hasil evaluasi dari kelima model yang diusulkan pada dataset yang sama.

#### **1.1.4. Research Gap**

Penelitian oleh Perdana et al., (2024) berkontribusi dalam deteksi promosi judi online di Twitter dengan pendekatan text mining klasik menggunakan kombinasi TF-IDF dan Random Forest. Namun, penelitian tersebut memiliki keterbatasan dalam proses pengumpulan data yang hanya menggunakan empat kata kunci eksplisit, yaitu “cuan”, “slot”, “untung”, dan “gacor”. Akibatnya, sebagian besar data yang terkumpul didominasi oleh promosi jenis permainan slot saja. Hal ini menghasilkan representasi yang sempit terhadap fenomena promosi judi online secara keseluruhan.

Berdasarkan analisis lanjutan menggunakan *word cloud*, terlihat bahwa istilah lain yang umum dalam dunia judi online seperti ‘scatter’, ‘return to player (rtp)’, ‘toto’, ‘macau’, dan lain-lain tidak muncul atau tidak ter-highlight secara signifikan. Padahal, dalam praktiknya, promosi judi online mencakup berbagai jenis permainan dan strategi promosi yang beragam. Berdasarkan beberapa sumber, kata kunci yang digunakan pada promosi judi online dapat dikelompokkan ke beberapa kategori utama yaitu istilah spesifik permainan, bahasa promosi dan insentif, eufemisme dan kata-kata halus, serta frasa ajakan terselubung (BagusEnrico, 2025; *The Ethics and Impact of Marketing Gambling: A Critical Examination - PR News*, n.d.; Torrance et al., 2021). Oleh karena itu, penelitian ini mengisi celah tersebut dengan membangun dataset yang lebih kaya dan representatif terhadap berbagai bentuk dan variasi promosi judi online, guna menghasilkan model deteksi yang lebih generalis, adaptif, dan relevan untuk kondisi nyata di berbagai platform sosial media.

Lebih lanjut, belum ditemukan penelitian yang secara eksplisit menerapkan dan membandingkan beragam pendekatan NLP seperti kombinasi *embedding* IndoBERTweet dengan klasifikator klasik, deep learning, hingga fine-tuning IndoBERTweet dalam konteks deteksi promosi judi online. Padahal, metode-metode tersebut telah terbukti efektif dalam tugas-tugas klasifikasi teks serupa seperti ujaran kebencian, sarkasme, dan berita palsu. Dengan demikian, penelitian ini tidak hanya memperluas cakupan representasi data, tetapi juga mengeksplorasi efektivitas berbagai pendekatan yang belum diterapkan secara khusus dalam domain ini.

### **1.2. Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini bertujuan untuk menjawab rumusan masalah berikut:

- Bagaimana efektivitas berbagai pendekatan NLP dalam mendeteksi promosi judi online di media sosial berbahasa Indonesia, khususnya yang menggunakan bahasa tidak baku, slang, dan konteks implisit?
- Apakah pendekatan kontekstual berbasis transformer (seperti IndoBERTweet) mampu mengungguli pendekatan leksikal dan pola lokal dalam tugas klasifikasi promosi judi online?
- Bagaimana performa masing-masing pendekatan Random Forest + TF-IDF, IndoBERTweet + BiLSTM, RF/SVM + IndoBERTweet, dan Fine-Tuned IndoBERTweet dalam mendeteksi berbagai bentuk promosi judi online berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score?

### **1.3. Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk mengembangkan dan mengevaluasi berbagai pendekatan NLP, termasuk model klasik, deep learning, hibrida, dan transformer (termasuk IndoBERTweet), dalam mendeteksi komentar promosi judi online berbahasa Indonesia di media sosial. Penelitian ini juga bertujuan mengatasi keterbatasan pendekatan sebelumnya yang hanya mengandalkan kata kunci eksplisit, dengan membangun dataset yang lebih beragam dan representatif, serta membandingkan performa masing-masing model berdasarkan akurasi, presisi, recall, dan F1-score dalam menghadapi bahasa tidak baku, slang, dan konteks implisit.

#### 1.4. Ruang Lingkup

Ruang lingkup dibatasi pada:

- a. Penelitian ini akan menguji dan membandingkan lima model yang merepresentasikan berbagai paradigma:
    - **Baseline Machine Learning:** Random Forest + TF-IDF.
    - **Hibrida Machine Learning + Transformer:**
      - Random Forest + IndoBERTweet (sebagai ekstraktor fitur).
      - SVM + IndoBERTweet (sebagai ekstraktor fitur).
    - **Deep Learning + Transformer:** IndoBERTweet + BiLSTM.
    - **State-of-the-Art Transformer:** IndoBERTweet (Fine-Tuned).
  - b. Input berupa dataset teks komentar sosial media yang terdiri atas 2 kolom “komentar” dan “label” dalam bentuk csv atau excel.
  - c. Data akan dikumpulkan dari berbagai platform media sosial yang sering menjadi tempat tersebarnya promosi judi online, seperti YouTube, Facebook, dan Threads, dengan menggunakan metode web scraping. Selain itu, data yang dikumpulkan akan dikombinasikan dengan dataset publik yang relevan dari platform seperti Kaggle dan Hugging Face. Adapun dataset publik yang digunakan antara lain:
    - **Dataset komentar siaran langsung Wayang Jogja Night Carnival #9**, yang berisi komentar dari live streaming dalam rangka HUT ke-268 Kota Yogyakarta. Dataset ini digunakan untuk mendeteksi promosi judi online dalam konteks interaksi langsung di media sosial (*Deteksi Judi Online*, n.d.).
    - **Dataset komentar YouTube**, yang telah dikumpulkan dan diberi label terkait indikasi promosi judi online (*Indirapravianti/Online\_gambling\_yt\_comments · Datasets at Hugging Face*, n.d.).
- Kedua dataset tersebut telah melalui proses pelabelan sebelumnya dan akan digunakan untuk memperkaya serta melengkapi data hasil scraping.
- d. Proses pencarian akan menggunakan strategi tiga lapis untuk cakupan yang komprehensif: (1) kata kunci eksplisit dari penelitian sebelumnya (Perdana et al., 2024), seperti 'slot' dan 'gacor'; (2) kata kunci baru yang diidentifikasi oleh penelitian tersebut sebagai temuan; dan (3) berbagai istilah dan slang kontemporer yang identik dengan judi online. Daftar kata kunci ini disusun dari hasil analisis berbagai sumber online, seperti forum, artikel glosarium, dan Google Trends yang membahas istilah terkait judi online (*Casino & Gambling Glossary - Glossary of Gambling Terms*, n.d.; *Essential Casino*

*Terms and Phrases*, n.d.; *Online Sports Betting Glossary*, n.d.; October 29 & Martinez, n.d.; Tadic, 2025). Tidak semua istilah akan digunakan, hanya yang paling relevan, kontekstual, dan sering muncul dalam promosi yang akan dipilih. Berikut ini kategorisasi dan daftar kata kuncinya:

Kategori/ Pola Linguistik	Kata Kunci	Deskripsi
Kata kunci spesifik game dan istilah terkait	“slot”, “toto”, “togel”, “poker”, “domino”, “macau”, “pola”	Menunjukkan jenis permainan judi atau istilah teknis yang hanya muncul di konteks judi.
Bahasa promosi dan berorientasi pada insentif	“free”, “bonus”, “jackpot”, “maxwin”, “promo”, “untung”, “withdraw”, “cuan”, “untung”, “gacor”	Mengandung ajakan terselubung dengan janji keuntungan besar secara finansial.
Eufemisme, bahasa gaul, dan bahasa tidak langsung	“link”, “situs”, “prediksi”, “mudah”, “aman”, “terbaik”, “resmi”, “terpercaya”, “terbaru”, “menang”, “rtp”, “scatter”	Kata-kata umum yang sering dipakai untuk menyamarkan aktivitas promosi judi.
Kata atau frasa ajakan secara langsung	“daftar”, “main”, “gabung”, “dapatkan”, “klaim”, “login”, “deposit”	Berisi instruksi atau ajakan eksplisit untuk melakukan aktivitas perjudian.

- e. Pengumpulan dataset juga diberi batasan sebagai berikut:
  - Hanya komentar publik yang ditulis dalam bahasa Indonesia.
  - Tidak mencakup gambar, video atau bentuk promosi non-teks.
- f. Target jumlah data minimal adalah 1000-3000 data.
- g. Proses pelabelan akan mengadopsi dan memperjelas aturan dari Perdana et al. (2024), dimana sebuah teks diklasifikasikan sebagai Promosi Judi Online jika memenuhi setidaknya salah satu dari kriteria berikut:
  - Mengandung kalimat yang secara terang-terangan mengajak atau mengundang untuk berpartisipasi dalam aktivitas judi online.

- Menyebutkan nama platform, menyertakan tautan (link) aktif, atau pola URL yang mengarah ke situs judi online.
  - Menunjukkan atau mengklaim hasil kemenangan, pencapaian (*maxwin*, *jackpot*), atau kemudahan penarikan dana (WD) dari aktivitas judi.
  - Memberikan arahan, tutorial, atau panduan tentang cara mendaftar, melakukan deposit, atau bermain di platform judi online.
- h. Dikarenakan tidak adanya pakar formal di bidang ini, proses pelabelan akan menggunakan pendekatan validasi silang antar anotator manusia untuk menjamin objektivitas dengan prosedur sebagai berikut.
- Setiap data akan dilabeli secara independen oleh minimal tiga orang anotator.
  - Untuk menjaga konsistensi dan mengurangi subjektivitas, setiap anotator akan berpegang pada seperangkat aturan pelabelan yang sama sebagaimana telah didefinisikan pada poin sebelumnya.
  - Label final untuk setiap data akan ditentukan melalui kesepakatan mayoritas (majority vote).
- i. Dataset mentah akan melalui proses pembersihan yang meliputi normalisasi kata slang dan tersamar (e.g., “s|ot,” “s1tus”), translasi emoji menjadi kata, lowercasing, dan tokenisasi.
- j. Dataset akan dievaluasi menggunakan metode *Stratified K-Fold Cross Validation* (misalnya K=5), untuk memastikan distribusi label yang seimbang di setiap fold serta hasil evaluasi yang lebih stabil dan adil.
- k. Tools yang digunakan (secara garis besar):
- *Python* – bahasa pemrograman utama.
  - *Hugging Face Transformers* – untuk *fine-tuning* model IndoBERTweet dan IndoBERT.
  - *scikit-learn* – untuk model baseline (TF-IDF + *Random Forest*).
  - *pandas/numpy* – manipulasi data.
  - *Matplotlib/Seaborn* – visualisasi hasil evaluasi.
  - *Jupyter Notebook / Google Colab / Kaggle* – untuk eksperimen dan dokumentasi awal.
- l. *Output* berupa hasil klasifikasi apakah sebuah komentar teks berisi indikasi promosi judi online atau bukan dengan format *true* atau *false*.

- m. Performa setiap model akan diukur dan dibandingkan menggunakan metrik standar: confusion matrix, accuracy, precision, recall, dan F1-score.
- n. Penelitian ini bersifat eksploratif dan terbatas pada tahap eksperimen; model yang dihasilkan tidak akan diintegrasikan ke sistem moderasi otomatis atau digunakan secara *real-time*.

## **1.5. Metodologi Penelitian**

### **1.5.1. Studi Literatur**

- 1.5.1.1. Studi mengenai karakteristik promosi judi online.
- 1.5.1.2. Studi mengenai metode machine learning (Random Forest, SVM)
- 1.5.1.3. Studi mengenai metode deep learning (BiLSTM).
- 1.5.1.4. Studi mengenai metode representasi teks (TF-IDF).
- 1.5.1.5. Studi mengenai model transformer (BERT, IndoBERT, IndoBERTweet).
- 1.5.1.6. Studi mengenai metrik evaluasi: Akurasi, Precision, Recall, F1-Score.

### **1.5.2. Pengumpulan Dataset dan *Preprocessing***

- 1.5.2.1. Mengambil komentar dari YouTube, Facebook, Threads via scraping.
- 1.5.2.2. Menambahkan dataset publik dari Kaggle & Hugging Face.
- 1.5.2.3. Melabeli data dengan kriteria promosi judi online (majority vote).
- 1.5.2.4. Preprocessing dengan normalisasi, translasi emoji, lowercasing, tokenisasi.
- 1.5.2.5. Analisis Word Cloud untuk melihat kata-kata dominan.
- 1.5.2.6. Split data untuk training dan testing.

### **1.5.3. Pembuatan dan Pelatihan Model**

- 1.5.3.1. Membangun 5 model klasifikasi : Fine-tuned IndoBERTweet, IndoBERTweet + BiLSTM, Random Forest + TF-IDF, Hybrid SVM + IndoBERTweet, Random Forest + IndoBERTweet.
- 1.5.3.2. Melatih model dengan Stratified K-Fold Cross Validation.

### **1.5.4. Pengujian dan Evaluasi**

- 1.5.4.1. Menguji performa semua model pada data uji.
- 1.5.4.2. Mengevaluasi performa berdasarkan metrik: Akurasi, Precision, Recall, dan F1-Score.
- 1.5.4.3. Membandingkan performa antar model untuk mengidentifikasi pendekatan paling efektif.

### **1.5.5. Analisis dan Pengambilan Kesimpulan**

1.5.5.1. Menganalisis perbedaan performa masing-masing pendekatan.

1.5.5.2. Menyimpulkan efektivitas metode yang digunakan dalam mendeteksi komentar promosi judi online.

#### **1.5.6. Pembuatan Laporan**

1.5.6.1. Menyusun laporan akhir berdasarkan hasil eksperimen, analisis, dan evaluasi.

### **1.6. Manfaat Penelitian**

- Memberikan wawasan baru tentang efektivitas berbagai pendekatan, termasuk machine learning, deep learning, dan transformer, dalam mendeteksi komentar promosi judi online di media sosial berbahasa Indonesia.
- Mendukung upaya moderasi konten digital dengan menghadirkan solusi otomatis yang mampu mengenali komentar promosi judi online yang bersifat singkat, tersamar, dan menggunakan bahasa tidak formal.
- Memberikan kontribusi terhadap pengembangan sistem deteksi konten ilegal berbasis AI yang lebih adaptif dan relevan dengan karakteristik bahasa Indonesia di media sosial.
- Memperkaya literatur ilmiah di bidang analisis teks sosial media dan penerapan model berbasis AI untuk isu-isu sosial yang berdampak terhadap keamanan digital dan literasi daring di Indonesia.

### **1.7. Sistematika Penulisan**

Penulisan skripsi ini dibagi menjadi beberapa bab yaitu:

#### **BAB I : PENDAHULUAN**

Pada bab ini berisikan latar belakang, rumusan masalah, ruang lingkup, tujuan penelitian, dan metodologi penelitian yang akan digunakan dalam skripsi ini.

#### **BAB II : LANDASAN TEORI**

Bab ini berisikan teori-teori yang digunakan untuk menjadi referensi dalam pembuatan skripsi dan diterapkan dalam skripsi ini.

#### **BAB III : METODE PENELITIAN**

Bab ini membahas analisis dari langkah-langkah apa saja yang akan dilakukan dalam membangun model.

#### **BAB IV : IMPLEMENTASI**



Bab ini berisikan penjelasan mengenai implementasi yang dilakukan berdasarkan pada rancangan yang tercantum pada bab III.

## BAB V : HASIL DAN PEMBAHASAN

Bab ini membahas tentang pengujian sistem yang telah diimplementasikan pada bab IV.

## BAB VI : KESIMPULAN DAN SARAN

Bab ini membahas tentang kesimpulan dari penelitian yang telah dilakukan serta saran bagi penelitian selanjutnya.

### 1.8.Jadwal Kegiatan

[illegible]

## **2. LANDASAN TEORI**

### **2.1. Tinjauan Pustaka**

#### **2.1.1. Promosi Judi Online**

Promosi judi online merupakan strategi pemasaran yang dilakukan oleh penyedia layanan perjudian digital untuk menarik partisipasi masyarakat dalam aktivitas perjudian. Seiring berkembangnya teknologi dan media sosial, promosi ini semakin mudah diakses dan menyebar luas melalui platform seperti *Twitter*, *Instagram*, dan *WhatsApp*. Fenomena ini menjadi perhatian serius karena dapat menimbulkan risiko sosial dan psikologis seperti kecanduan, kerugian ekonomi, serta pelanggaran hukum. Menurut Akbar et al. (2024), terpaan iklan judi online di media sosial berkontribusi signifikan terhadap peningkatan minat masyarakat, khususnya generasi muda, dalam mencoba judi online. Ciri khas promosi ini meliputi penggunaan bahasa yang positif, menonjolkan keuntungan cepat, akses mudah, serta iming-iming hadiah besar. Selain itu, konten promosi seringkali melibatkan *influencer* atau figur publik untuk menarik kepercayaan calon pemain (Angellina & Prasetyo, 2024).

#### **2.1.2. Random Forest**

Random Forest adalah algoritma ensemble berbasis pohon keputusan yang membangun banyak decision tree dari sampel acak (bootstrap) dan menggabungkan prediksinya melalui voting mayoritas (*What Is Random Forest?*, 2021). Dalam klasifikasi teks, model ini bekerja setelah teks diubah ke bentuk numerik seperti TF-IDF. Random Forest dikenal kuat terhadap overfitting, stabil, dan memberikan performa tinggi, meski membutuhkan lebih banyak memori dan sulit diinterpretasi (AIML.com, 2022).

#### **2.1.3. Support Vector Machine (SVM)**

Support Vector Machine (SVM) adalah algoritma klasifikasi yang mencari hyperplane optimal untuk memisahkan dua kelas data. SVM sangat cocok untuk data berdimensi tinggi seperti teks dan hanya menggunakan beberapa support vector, sehingga efisien dalam memori. Dengan kernel trick, SVM juga dapat memproyeksikan data ke ruang dimensi lebih tinggi agar dapat dipisahkan secara linear. Namun, waktu pelatihan dapat lama untuk dataset besar dan hasilnya sulit diinterpretasi (AIML.com, 2022; Editorial, 2022).

#### **2.1.4. TF-IDF (Term Frequency-Inverse Document Frequency)**

TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik pembobotan kata yang digunakan untuk merepresentasikan dokumen dalam bentuk vektor. TF mengukur frekuensi kata dalam dokumen, sedangkan IDF mengukur kelangkaan kata di seluruh korpus. TF-IDF menekankan kata-kata yang penting dan jarang, namun tidak memahami konteks atau urutan kata, serta menghasilkan matriks yang besar dan sparse (Ninja, 2023).

#### **2.1.5. Bidirectional Long Short Term Memory (BiLSTM)**

Arsitektur Bidirectional Long Short-Term Memory (BiLSTM) merupakan pengembangan dari RNN yang dirancang untuk mengatasi keterbatasan dalam memproses data sekuensial, khususnya masalah *vanishing gradient* yang menghambat pembelajaran dependensi jangka panjang. LSTM mengatasinya melalui mekanisme *cell state* dan *gating* (forget, input, output), yang memungkinkan model secara selektif menyimpan atau melupakan informasi (*What Is Long Short-Term Memory (LSTM)?*, n.d.). BiLSTM menyempurnakan ini dengan membaca urutan teks dari dua arah (maju dan mundur), sehingga mampu membangun representasi kontekstual yang lebih menyeluruh (*Papers with Code - BiLSTM Explained*, n.d.). Keunggulan ini menjadikannya efektif dalam memahami struktur kalimat dan konteks implisit dalam tugas klasifikasi. Meskipun transformer kini mendominasi banyak aplikasi NLP, BiLSTM tetap relevan, terutama dalam kondisi dengan data terbatas atau keterbatasan komputasi.

#### **2.1.6. Bidirectional Encoder Representations from Transformers (BERT)**

*Bidirectional Encoder Representations from Transformers* (BERT) adalah model bahasa berbasis arsitektur *transformer* yang dikembangkan oleh Google (Devlin et al., 2019). Keunggulannya terletak pada kemampuan memahami konteks kata secara dua arah melalui dua tugas utama: *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Struktur BERT menggunakan encoder dengan self-attention, menjadikannya efektif untuk berbagai tugas NLP seperti klasifikasi teks dan analisis sentimen. Meski tidak digunakan langsung dalam penelitian ini, BERT menjadi dasar dari model seperti IndoBERT dan IndoBERTweet yang lebih sesuai untuk bahasa Indonesia.

##### **2.1.6.1. IndoBERT**

IndoBERT adalah model bahasa pra-latih berbasis BERT yang dikembangkan oleh Koto et al. (2020) khusus untuk Bahasa Indonesia. Model ini dilatih menggunakan lebih dari 200

juta kata dari berbagai sumber seperti berita, Wikipedia, dan dokumen publik, dengan metode *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP). Dibandingkan mBERT, IndoBERT menunjukkan performa lebih baik dan konsisten dalam benchmark IndoLEM untuk tugas-tugas NLP seperti klasifikasi sentimen, NER, dan *POS tagging*, menegaskan keunggulan model monolingual untuk bahasa lokal.

#### 2.1.6.2. IndoBERTweet

IndoBERTweet adalah model bahasa pra-latih yang secara khusus dikembangkan untuk memproses teks dari media sosial *Twitter* berbahasa Indonesia (Koto et al., 2021). Model ini menggunakan arsitektur BERT dasar dan dilatih pada 26 juta *tweet* Indonesia yang telah diproses melalui tokenisasi khusus, termasuk penggantian mention dan URL dengan token khusus, serta konversi emoji menjadi teks. Salah satu inovasi penting dalam pengembangan IndoBERTweet adalah inisialisasi kosakata domain-spesifik, di mana token baru dari *Twitter* yang tidak dikenal oleh IndoBERT diinisialisasi menggunakan rata-rata vektor dari subword-nya berdasarkan embedding IndoBERT sebelumnya. Rumus perhitungannya adalah:

$$E_{\text{new}}(x) = \frac{1}{|T(x)|} \sum_{t \in T(x)} E(t)$$

Di mana  $T(x)$  adalah sekumpulan token *subword* dari kata baru  $x$ , dan  $E(t)$  adalah embedding dari masing-masing *subword*. Ini memastikan bahwa kata-kata baru yang khas di media sosial tetap memiliki representasi vektor yang relevan secara semantik. Selain itu, IndoBERTweet menggunakan teknik *Masked Language Modeling* (MLM) untuk melatih model memahami konteks dua arah dari teks. Rumus loss untuk MLM didefinisikan sebagai:

$$L_{\text{MLM}} = - \sum_{i \in M} \log P(x_i | x_{\setminus M})$$

Di mana  $M$  merupakan indeks token yang dimasking, dan  $x_{\setminus M}$  adalah konteks token lain di sekitarnya. Model ini telah terbukti unggul dalam berbagai tugas NLP berbasis *Twitter* seperti analisis sentimen, deteksi ujaran kebencian, dan klasifikasi emosi dibandingkan dengan model *baseline* seperti IndoBERT dan mBERT.

#### 2.1.7. Metrik Evaluasi

Evaluasi performa model klasifikasi merupakan langkah penting dalam pengembangan model *machine learning*. Untuk menilai seberapa baik model memprediksi data, digunakan beberapa metrik evaluasi seperti *confusion matrix*, akurasi, presisi, *recall*, dan F1-score.

Metrik-metrik ini memungkinkan peneliti untuk memahami kelebihan dan kelemahan model dalam berbagai aspek, terutama saat menghadapi data yang tidak seimbang.

#### 2.1.7.1. Confusion Matrix

Confusion matrix adalah representasi visual dari performa model klasifikasi dengan menampilkan jumlah prediksi benar dan salah untuk setiap kelas (Géron, 2022). Matrix ini terdiri dari empat komponen utama:

- *True Positive* (TP): jumlah data *positif* yang diprediksi benar oleh model.
- *False Positive* (FP): jumlah data *negatif* yang diprediksi sebagai *positif*.
- *True Negative* (TN): jumlah data *negatif* yang diprediksi benar sebagai *negatif*.
- *False Negative* (FN): jumlah data *positif* yang diprediksi sebagai *negatif*.

#### 2.1.7.2. Akurasi

Akurasi mengukur proporsi prediksi yang benar dari seluruh prediksi yang dilakukan oleh model (Géron, 2022). Nilai akurasi yang tinggi menunjukkan bahwa model dapat melakukan prediksi yang benar pada sebagian besar data. Namun, pada data yang tidak seimbang (*imbalanced*), akurasi bisa menyesatkan karena model bisa saja hanya menebak kelas mayoritas dan tetap terlihat “baik”. Nilai akurasi dihitung dengan rumus:

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN}$$

#### 2.1.7.3. Presisi

Presisi mengukur seberapa akurat model dalam memprediksi kelas positif, yaitu proporsi prediksi positif yang benar-benar positif (Géron, 2022). Presisi yang tinggi menunjukkan bahwa sebagian besar prediksi positif benar-benar relevan. Presisi rendah berarti model terlalu banyak menghasilkan *false positive*. Presisi dapat dihitung dengan rumus berikut:

$$\text{Presisi} = \frac{TP}{TP + FP}$$

#### 2.1.7.4. Recall

*Recall* atau sensitivitas menunjukkan seberapa baik model dalam menangkap semua data positif (Géron, 2022). *Recall* tinggi berarti model mampu menemukan sebagian besar data yang benar-benar positif. *Recall* rendah berarti model melewatkan banyak data positif

(banyak *false negative*), yang bisa sangat berbahaya dalam konteks seperti deteksi penipuan, penyakit, atau promosi judi online. Nilai *recall* dihitung dengan rumus:

$$\text{Recall} = \frac{TP}{TP + FN}$$

#### 2.1.7.5. F1-Score

Dalam konteks klasifikasi biner, *precision* dan *recall* merupakan dua metrik penting untuk mengevaluasi performa model. Namun, dalam banyak kasus, kedua metrik ini tidak cukup jika dilihat secara terpisah. Oleh karena itu, sering digunakan metrik gabungan yang disebut *F1-Score*, yaitu rata-rata harmonik dari *precision* dan *recall* (Géron, 2022).

$$\text{F1-Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

Berbeda dengan rata-rata aritmatika, rata-rata harmonik memberikan bobot lebih besar terhadap nilai yang rendah (Géron, 2022). Artinya, *F1 Score* akan menghasilkan nilai tinggi hanya jika *precision* dan *recall* sama-sama tinggi. Jika salah satu rendah, maka *F1 Score* juga akan rendah. Hal ini membuat *F1 Score* sangat berguna ketika diperlukan keseimbangan antara *precision* dan *recall*.

## 2.2. Tinjauan Studi

### 2.2.1. Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm

Penelitian oleh Perdana et al. (2024) mencoba menjawab tantangan dalam mendeteksi promosi judi online di Twitter berbahasa Indonesia, yang masih jarang diteliti. Menggunakan 6.038 tweet, studi ini menguji efektivitas tiga algoritma klasifikasi—Random Forest, Logistic Regression, dan CNN—dengan tiga metode representasi teks: TF-IDF, Word2Vec, dan GloVe. Hasil terbaik didapat dari kombinasi TF-IDF dan Random Forest, yang mencapai *precision* 0,966 dan *recall* 0,958. Beberapa kata kunci yang sering muncul dalam konten promosi judi antara lain *link*, *situs*, *prediksi*, *jackpot*, *maxwin*, dan *togel*. Temuan ini menunjukkan potensi besar dalam membantu upaya keamanan siber dan pemberantasan konten negatif di media sosial.

### **2.2.2. Fine-Tuning BERT-Based Models for Negative Content Identification on Indonesian Tweets**

Penelitian oleh Hidayatullah et al. (2023) menyoroti pentingnya moderasi konten di media sosial, khususnya Twitter, yang kini dipenuhi oleh konten negatif seperti ujaran kasar, konten seksual eksplisit, dan ujaran kebencian. Meskipun isu ini krusial, kajian terhadap konten berbahasa Indonesia masih terbatas. Untuk mengisi kekosongan tersebut, studi ini mengevaluasi efektivitas model BERT, khususnya IndoBERTweet, dalam mengklasifikasi tweet Indonesia ke dalam kategori abusive, dewasa, dan netral. Hasilnya menunjukkan bahwa IndoBERTweet memberikan performa terbaik, dengan precision sebesar 97.03, recall 96.88, dan macro F1 score 96.94. Temuan ini menunjukkan potensi besar model berbasis BERT dalam mendukung terciptanya lingkungan daring yang lebih sehat dan aman.

### **2.2.3. Implementation of IndoBERT in Sarcasm Detection using Random Forest Towards Sentiment Analysis**

Penelitian oleh Sibarani et al., (2025) berusaha mengatasi tantangan deteksi sarkasme dalam teks bahasa Indonesia menggunakan Random Forest Classifier (RFC) yang dipadukan dengan IndoBERT. Evaluasi dengan 10-fold dan 5-fold cross-validation menunjukkan RFC tanpa IndoBERT memiliki performa sekitar 78-79% pada akurasi, presisi, recall, dan F1-score. Setelah integrasi IndoBERT, semua metrik meningkat signifikan, bahkan mencapai 97,24% pada 5-fold cross-validation. Hasil ini menegaskan efektivitas IndoBERT dalam memperkuat deteksi sarkasme sesuai konteks bahasa Indonesia untuk meningkatkan akurasi analisis sentimen di media sosial.

### **2.2.4. Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM**

Penelitian Kusuma & Chowanda (2023) menggabungkan model IndoBERTweet dan BiLSTM untuk mendeteksi ujaran kebencian dalam bahasa Indonesia. IndoBERTweet dipilih karena dirancang khusus untuk teks Twitter yang cenderung informal, sementara BiLSTM digunakan untuk memahami konteks kata dari dua arah. Model dilatih menggunakan data berlabel dari Twitter dan berhasil mengklasifikasikan tweet ke dalam dua kelas: hate speech dan non-hate speech. Hasilnya menunjukkan akurasi tinggi sebesar 93,7%, melampaui metode sebelumnya, sehingga membuktikan efektivitas pendekatan ini dalam menangani ujaran kebencian di media sosial.

#### **2.2.5. Hate Speech Detection and Online Public Opinion Regulation Using Support Vector Machine Algorithm: Application and Impact on Social Media**

Penelitian Li & Li, (2025) menggunakan SVM dengan teknik kernel untuk mendeteksi ujaran kebencian di media sosial yang kompleks dan tersirat, seperti sarkasme dan variasi ejaan. Data dikumpulkan lewat API dan crawler dengan autentikasi OAuth2.0, kemudian diproses dengan pembersihan, Word2Vec Skip-gram, TF-IDF, serta ekstraksi fitur yang menggabungkan analisis sentimen berbasis leksikon dan BERT. Evaluasi pada dua dataset menunjukkan akurasi 90,42% dan 92,84%, recall 88,06% dan 90,79%, serta waktu inferensi cepat di bawah 4 ms. Hasil ini membuktikan efektivitas SVM dalam mendeteksi ujaran kebencian implisit secara akurat dan efisien untuk pemantauan real-time dan keamanan daring.



## DAFTAR PUSTAKA

- AIML.com. (2022, June 11). What are the advantages and disadvantages of Random Forest?  
*AIML.Com*. <https://aiml.com/what-are-the-advantages-and-disadvantages-of-random-forest/>
- Akbar, M. Y., Pradekso, T., & Ulfa, N. S. (2024). PENGARUH TERPAAN IKLAN JUDI ONLINE DI MEDIA SOSIAL, TINGKAT PENGAWASAN ORANG TUA, DAN INTENSITAS KOMUNIKASI PEER GROUP TERHADAP MINAT BERMAIN JUDI ONLINE. *Interaksi Online*, 12(4), Article 4.
- Angellina, P., & Prasetyo, B. (2024). Pertanggungjawaban Pidana Terhadap Pelaku yang Mempromosikan Judi Online. *Ranah Research : Journal of Multidisciplinary Research and Development*, 7(2), 946–952. <https://doi.org/10.38035/rrj.v7i2.1395>
- Apresiasi Laporan Masyarakat, Komdigi Tangani 1,3 Juta Konten Pornografi dan Judi Online*. (2025, March 10). <https://www.komdigi.go.id/berita/siaran-pers/detail/apresiasi-laporan-masyarakat-komdigi-tangani-13-juta-konten-pornografi-dan-judi-online>
- BagusEnrico. (2025, June 10). Online Gambling in Indonesia: Regulatory Landscape, Current Developments, and Economic Perspective on a Growing Digital Threat. *Bagus Enrico & Partners*. <https://bagusenrico.com/2025/06/10/online-gambling-in-indonesia-regulatory-landscape-current-developments-and-economic-perspective-on-a-growing-digital-threat/>
- Casino & Gambling Glossary—Glossary of Gambling Terms*. (n.d.). Casino.Org. Retrieved June 19, 2025, from <https://www.casino.org/glossary/>
- Deteksi judi online*. (n.d.). Retrieved June 9, 2025, from <https://www.kaggle.com/datasets/yaemico/deteksi-judi-online>

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Editorial. (2022, September 10). Pros and cons of Support Vector Machine (SVM). *RoboticsBiz*. <https://roboticsbiz.com/pros-and-cons-of-support-vector-machine-svm/>
- Essential Casino Terms and Phrases*. (n.d.). Retrieved June 19, 2025, from <https://sigma.world/play/blog/casino-terms/>
- Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (Third). O'Reilly Media.
- Handoko, Junadhi, Triyani Arita Fitri, & Agustin. (2025). Optimization of Deep Learning with FastText for Sentiment Analysis of the SIREKAP 2024 Application. *The Indonesian Journal of Computer Science*, 14(2). <https://doi.org/10.33022/ijcs.v14i2.4809>
- Hidayatullah, A. F., Kalinaki, K., Aslam, M. M., Zakari, R. Y., & Shafik, W. (2023). Fine-Tuning BERT-Based Models for Negative Content Identification on Indonesian Tweets. *2023 8th International Conference on Information Technology and Digital Applications (ICITDA)*, 1–6. <https://doi.org/10.1109/ICITDA60835.2023.10427046>
- Huda, A. M., Shidik, G. F., & Praskatama, V. (2024). COMPARATIVE ANALYSIS OF LSTM, BILSTM, GRU, CNN, AND RNN FOR DEPRESSION DETECTION IN SOCIAL MEDIA. *Jurnal Teknik Informatika (Jutif)*, 5(6), Article 6. <https://doi.org/10.52436/1.jutif.2024.5.6.4060>
- Indirapravianti/online\_gambling\_yt\_comments · Datasets at Hugging Face*. (n.d.). Retrieved June 9, 2025, from [https://huggingface.co/datasets/indirapravianti/online\\_gambling\\_yt\\_comments](https://huggingface.co/datasets/indirapravianti/online_gambling_yt_comments)
- Kaware, P. D. (2025). Indo-HateSpeech Analysis: A Multi-Level Hate Speech Classification Framework Using BERT Features and Machine Learning Models. *Advances in Nonlinear Variational Inequalities*, 28(5s), Article 5s. <https://doi.org/10.52783/anvi.v28.3912>

- Khaire, S., Mane, S., Milake, S., & Bagade, D. A. (2025). *Social Media Fake News Detection Using DistilBERT Algorithm And SVM Model*. 12(01).
- Koto, F., Lau, J. H., & Baldwin, T. (2021). *IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization* (arXiv:2109.04607). arXiv. <https://doi.org/10.48550/arXiv.2109.04607>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 757–770). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Kusuma, J. F., & Chowanda, A. (2023). Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM on Twitter. *JOIV : International Journal on Informatics Visualization*, 7(3), 773–780. <https://doi.org/10.30630/joiv.7.3.1035>
- Li, S., & Li, Z. (2025). Hate Speech Detection and Online Public Opinion Regulation Using Support Vector Machine Algorithm: Application and Impact on Social Media. *Information*, 16(5), Article 5. <https://doi.org/10.3390/info16050344>
- Mufrida, I. E. (2024, March 30). *Orang Indonesia Paling Sering Habiskan Waktu untuk Main Sosial Media* [Article]. GoodStats. <https://goodstats.id/article/orang-indonesia-paling-sering-habiskan-waktu-untuk-main-sosial-media-nETfh>
- Nabiilah, G. Z., Prasetyo, S. Y., Izdiyar, Z. N., & Girsang, A. S. (2023). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, 216, 714–721. <https://doi.org/10.1016/j.procs.2022.12.188>
- Ninja, N. (2023, June 30). TF-IDF: Weighing Importance in Text. *Let's Data Science*. <https://letsdatascience.com/tf-idf/>

Nuha, U., & Lin, C.-H. (2025). Investigating the Contribution of Structural and Contextual Features for Spam Detection. *Tehnički Vjesnik*, 32(3), 958–965.

<https://doi.org/10.17559/TV-20240917001995>

October 29, B. T. U., & Martinez, 2024Fact Checked ByLogan. (n.d.). *Online Casino Glossary of Terms*. Action Network. Retrieved June 19, 2025, from

<https://www.actionnetwork.com/casino/online-casino-glossary>

*Online Sports Betting Glossary: Understanding the Language of the Industry*. (n.d.). Retrieved June 19, 2025, from <https://www.watechnology.com/news/online-sports-betting-glossary-understanding-the-language-of-the-industry>

*Papers with Code—BiLSTM Explained*. (n.d.). Retrieved June 21, 2025, from

<https://paperswithcode.com/method/bilstm>

Perdana, R. B., -, A., Budi, I., Santoso, A. B., Ramadiah, A., & Putra, P. K. (2024). Detecting Online Gambling Promotions on Indonesian Twitter Using Text Mining Algorithm.

*International Journal of Advanced Computer Science and Applications*, 15(8).

<https://doi.org/10.14569/ijacsa.2024.0150893>

Rauf, T., & US. (2025, January 7). *(SIARAN PERS) Sejak Awal Januari 2025 Kemkomdigi Telah Menindak 43 Ribu Konten Judol*.

[https://www.subang.go.id/public/index.php/berita/siaran-pers-sejak-awal-januari-2025-kemkomdigi-telah-menindak-43-ribu-konten-judol?utm\\_source=chatgpt.com](https://www.subang.go.id/public/index.php/berita/siaran-pers-sejak-awal-januari-2025-kemkomdigi-telah-menindak-43-ribu-konten-judol?utm_source=chatgpt.com)

Selpiyani, Y. (2024). *RRI.co.id—Bahaya dan Dampak Judi Online: Ancaman yang Harus*

*Diwaspadai*. <https://www.rri.co.id/nasional/783149/bahaya-dan-dampak-judi-online-ancaman-yang-harus-diwaspadai>

Setiadi, B. A. (2024). *Menkominfo: Judi Online Ganggu Fondasi Ekonomi Negara*.

<https://www.komdigi.go.id/berita/siaran-pers/detail/menkominfo-judi-online-ganggu-fondasi-ekonomi-negara>

- Sibarani, S. A. B., Purba, R., & Limbong, R. P. (2025). Implementation of IndoBERT in Sarcasm Detection using Random Forest Towards Sentiment Analysis. *Building of Informatics, Technology and Science (BITS)*, 6(4), Article 4. <https://doi.org/10.47065/bits.v6i4.5801>
- Tadic, L. (2025, May 7). Slot Glossary: Slot Machine Terminology & Slot Terms Explained. *SigmaPlay*. <https://sigma.world/play/blog/slots-glossary/>
- The Ethics and Impact of Marketing Gambling: A Critical Examination—PR News*. (n.d.). Retrieved June 17, 2025, from <https://everything-pr.com/the-ethics-and-impact-of-marketing-gambling-a-critical-examination/>
- Torrance, J., John, B., Greville, J., O’Hanrahan, M., Davies, N., & Roderique-Davies, G. (2021). Emergent gambling advertising; a rapid review of marketing content, delivery and structural features. *BMC Public Health*, 21, 718. <https://doi.org/10.1186/s12889-021-10805-w>
- What Is Long Short-Term Memory (LSTM)?* (n.d.). Retrieved June 21, 2025, from <https://www.mathworks.com/discovery/lstm.html>
- What Is Random Forest? | IBM*. (2021, October 20). <https://www.ibm.com/think/topics/random-forest>