Jurnal Sifo Mikroskil (JSM)
Volume 27, No 2, Oktober 2026 – Hal 235 - 250
DOI: https://doi.org/10.55601/jsm.v27i2.2060

e-ISSN: 2622-8130
ISSN: 1412-0100

# Implementasi Aplikasi Deteksi Komentar Spam Iklan Judi Online pada Youtube Berbahasa Indonesia Menggunakan Indobert

**Rohid Zaidan Nasution[1], Syanti Irviantina[2], Erlina Halim[3]**
[1,2,3]Universitas Mikroskil, Jl. Thamrin No.124 Medan, (061) 4573767
[1,2,3]Fakultas Informatika, Teknik Informatika, Universitas Mikroskil, Medan
e-mail: [1]221111415@alumni.mikroskil.ac.id, [2]syanti@mikroskil.ac.id, [3]erlina.halim@mikroskil.ac.id

*Dikirim:* ***06-07-2026*** | *Diterima:* ***14-10-2026*** | *Diterbitkan:* ***31-10-2026***

**Abstrak**

Maraknya komentar spam judi online di YouTube yang memanfaatkan teknik penyamaran (*obfuscation*) dan bahasa gaul (*slang*) menjadi tantangan serius yang sulit diatasi oleh sistem filtrasi otomatis berbasis kata kunci konvensional. Penelitian ini bertujuan mengimplementasikan model *Deep Learning* IndoBERT untuk mendeteksi komentar spam iklan judi online berbahasa Indonesia secara otomatis pada platform YouTube dan mengintegrasikannya ke dalam aplikasi peramban *Chrome Extension*. Model dilatih menggunakan teknik *fine-tuning* dengan *learning rate* 2e-5, *batch size* 16, dan 3 *epoch* pada dataset komentar YouTube. Hasil evaluasi eksperimen menunjukkan bahwa model IndoBERT sangat efektif menangani ketidakseimbangan kelas (*imbalanced dataset*), mencapai nilai Akurasi sebesar 95,05%, Presisi 91%, dan *F1-Score* 90% pada kelas spam. Tingginya nilai presisi membuktikan bahwa sistem memiliki tingkat kesalahan *False Positive* yang sangat rendah, sehingga aman meminimalkan risiko salah blokir komentar normal. Sistem diimplementasikan dalam arsitektur *Client-Server*, di mana ekstensi peramban memanipulasi *Document Object Model* (DOM) halaman YouTube secara *real-time* melalui pilihan tindakan *blurring*, *hide*, dan pemberian label *warning* berdasarkan prediksi API server berbasis Flask. Integrasi ini terbukti menjadi solusi teknis yang praktis dan efektif untuk membantu pengguna menciptakan lingkungan digital yang bersih dari promosi judi online.

**Kata kunci**: Deteksi_Spam, Judi_Online, IndoBERT, Chrome_Extension, Deep_Learning

***Abstract***

*The proliferation of online gambling spam comments on YouTube utilizing obfuscation techniques and slang poses a serious challenge that conventional keyword-based automatic filters fail to address effectively. This study aims to implement the IndoBERT deep learning model for automatically detecting Indonesian online gambling spam comments on YouTube and integrating it into a Chrome Extension application. The model was trained using fine-tuning techniques with a learning rate of 2e-5, a batch size of 16, and 3 epochs on a YouTube comment dataset. Experimental evaluation results demonstrate that the IndoBERT model effectively handles class imbalance, achieving an Accuracy of 95.05%, Precision of 91%, and an F1-Score of 90% for the spam class. The high precision score indicates a low False Positive rate, ensuring normal comments are safely preserved. The system was successfully deployed using a Client-Server architecture, where the browser extension dynamically manipulates YouTube's Document Object Model (DOM) in real-time through blurring, hiding, and warning actions based on predictions from a Flask-based API server. This integration provides a practical and effective technical solution to help users maintain a clean digital environment free from online gambling promotions.*

***Keywords****: Spam_Detection, Online_Gambling, IndoBERT, Chrome_Extension, Deep_Learning*

## 1. PENDAHULUAN

Perkembangan teknologi digital telah mempengaruhi cara masyarakat berinteraksi di media sosial, salah satunya melalui platform YouTube yang menyediakan ruang komunikasi interaktif skala besar. Namun, keterbukaan ini juga menghadirkan tantangan baru berupa maraknya komentar spam yang mengandung promosi aktivitas ilegal seperti judi online. Fenomena tersebut tidak hanya mengganggu kenyamanan pengguna, tetapi juga merusak citra platform digital [1], [2].

Selain memberikan gangguan visual, maraknya komentar spam judi online juga menimbulkan tantangan serius karena sering menggunakan teknik *obfuscation* (penyamaran karakter) dan bahasa tidak baku untuk menghindari filter kata kunci standar. Sistem moderasi bawaan YouTube saat ini belum sepenuhnya efektif dalam mendeteksi variasi penulisan tersebut. Kegagalan filtrasi otomatis tanpa diimbangi dengan solusi praktis di sisi pengguna dapat meningkatkan paparan masyarakat terhadap risiko konten ilegal. Oleh karena itu, diperlukan sebuah mekanisme deteksi cerdas yang adaptif terhadap konteks lokal Bahasa Indonesia guna memastikan ruang digital yang aman dan bersih [3].

Untuk menganalisis dan mengatasi masalah klasifikasi teks yang kompleks ini, penelitian ini menggunakan model *deep learning* berbasis *Transformer* yaitu IndoBERT. Model ini memiliki keunggulan utama melalui mekanisme *Bidirectional Attention* yang memungkinkannya memahami konteks kalimat secara utuh dari dua arah, serta proses *pre-training* menggunakan *corpus* besar Bahasa Indonesia. Karakteristik tersebut membuat IndoBERT sangat sensitif terhadap bahasa tidak baku, singkatan, dan istilah khas media sosial (seperti "slot gacor" atau "togel") yang sering digunakan dalam komparasi spam judi online [4], [5].

Tujuan penelitian ini adalah untuk mengimplementasikan model IndoBERT dalam sistem deteksi komentar spam iklan judi online berbahasa Indonesia pada platform YouTube dan mengembangkannya dalam bentuk aplikasi *Chrome Extension*. Penelitian ini bertujuan untuk menjembatani kesenjangan antara hasil penelitian akademik yang masih berbasis *Command Line Interface* (CLI) dan kebutuhan nyata masyarakat akan sistem filtrasi yang praktis, aplikatif, serta dapat bekerja secara *real-time* di sisi *client*.

Penelitian terdahulu membuktikan bahwa model Transformer berbasis IndoBERT mampu mendeteksi teks judi online dengan akurasi dan keseimbangan metrik yang sangat tinggi [4]. Meskipun metode konvensional seperti *Support Vector Machine* (SVM) [1] maupun kombinasi *Regex* dan *Fuzzy Matching* [2] telah diterapkan, pendekatan tersebut masih memiliki keterbatasan dalam menangani fleksibilitas ejaan baru. Selain itu, sebagian besar implementasi yang ada umumnya masih terisolasi dalam lingkungan eksperimental berbasis *Command Line Interface* (CLI) tanpa antarmuka yang praktis bagi pengguna akhir (*end-user*). Untuk menjembatani celah tersebut, penelitian ini mengintegrasikan model *fine-tuned* IndoBERT ke dalam aplikasi *Chrome Extension* berbasis arsitektur *Client-Server* (Flask API) [6] . Kontribusi utama dari pendekatan ini adalah kemampuannya memanipulasi *Document Object Model* (DOM) YouTube secara *real-time* melalui skema proteksi visual (*blur*, *hide*, dan *warn*) yang dapat disesuaikan dengan preferensi pengguna.

## 2. TINJAUAN PUSTAKA

### 2.1 Deteksi Komentar Spam dan Natural Language Processing (NLP)

Deteksi komentar spam merupakan proses identifikasi pesan atau tanggapan yang bersifat tidak relevan, berbentuk promosi, mengandung tautan berbahaya, atau dilakukan secara otomatis oleh *bot* [7]. Pada platform YouTube, komentar spam sering kali dimodifikasi dengan penggunaan karakter non-standar atau teknik *obfuscation* guna menghindari sistem penyaringan kata kunci konvensional [8].

Dalam menangani masalah ini, *Natural Language Processing* (NLP) berperan penting sebagai cabang kecerdasan buatan yang menjembatani komunikasi antara bahasa manusia dan pemrosesan

komputer [9]. NLP melakukan transformasi teks ke dalam representasi numerik melalui tahapan *text preprocessing* seperti *case folding*, *tokenization*, *stopword removal*, dan *stemming* [10]. Hasil pra-pemrosesan tersebut selanjutnya diolah menggunakan metode ekstraksi fitur seperti *Bag-of-Words* (BoW), TF-IDF, atau *word embeddings* untuk keperluan klasifikasi teks menjadi kategori *spam* dan *non-spam* [11], [12].

### 2.2 Model Machine Learning untuk Deteksi Spam

*Machine Learning* dalam tugas klasifikasi teks berfokus pada algoritma yang memungkinkan komputer mengenali pola secara mandiri berdasarkan data pelatihan [3]. Secara umum, pendekatan pemecahan masalah dengan *Machine Learning* terbagi menjadi tiga kategori utama [13]:

1. *Supervised Learning*: Pembelajaran terarah menggunakan data training yang telah memiliki label kelas, seperti algoritma *Logistic Regression*, SVM, *Random Forest*, dan jaringan saraf tiruan.
2. *Unsupervised Learning*: Pembelajaran tanpa petunjuk atau label di mana komputer mencari pola tersembunyi secara mandiri, yang umumnya dibagi menjadi tugas asosiasi dan *clustering*.
3. *Reinforcement Learning*: Proses pembelajaran interaktif berbasis lingkungan di mana model mengambil keputusan spesifik berdasarkan skema penghargaan (*reward*) dan hukuman (*punishment*).

Model tradisional seperti SVM atau Naïve Bayes memiliki keunggulan dari efisiensi komputasi, namun kinerjanya menurun saat menghadapi teks dengan variasi ejaan tidak baku [14]. Untuk mengatasi batasan tersebut, pendekatan *Deep Learning* seperti BiLSTM diterapkan karena mampu membaca konteks kalimat dari dua arah [15]. Perkembangan terkini menunjukkan tren peralihan ke arsitektur berbasis Transformer yang memanfaatkan mekanisme *self-attention* untuk memproses hubungan antar-kata di seluruh kalimat secara paralel [16].

### 2.3 IndoBERT

IndoBERT merupakan adaptasi dari arsitektur *Transformer* berbasis *encoder-only* (BERT) yang secara khusus dilatih menggunakan korpus besar berbahasa Indonesia [17]. Model ini dirancang untuk memahami morfologi, afiksasi, serta ragam kosakata informal yang sering muncul pada media sosial Indonesia [18].

Arsitektur dasar IndoBERT terdiri atas beberapa lapisan *Transformer* dengan tiga komponen utama di setiap lapisannya [19]:

1. *Self-Attention Layer*: Menggunakan mekanisme *Multi-Head Self-Attention* untuk menentukan tingkat keterkaitan antar-kata dan menangani makna kata yang ambigu berdasarkan konteksnya.
2. *Feed-Forward Neural Network*: Jaringan *fully-connected* yang berfungsi memproyeksikan representasi hasil *attention* ke dimensi yang lebih tinggi demi memperkaya kapasitas komputasi model.
3. *Normalization* dan *Residual Connection*: Komponen jalan pintas (*shortcut*) untuk memperlancar aliran gradien dan menstabilkan distribusi aktivasi antar-lapisan guna mencegah *vanishing gradient*.

Sebelum diproses, teks dikonversi menjadi representasi numerik melalui proses *embedding* yang menggabungkan *Token Embedding* (berbasis *WordPiece Tokenization*), *Segment Embedding* (pembeda batas kalimat), dan *Positional Embedding* (informasi urutan kata) [18]. Pada tahap *pre-training*, IndoBERT dilatih melalui tugas *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP) [20]. Selanjutnya, melalui konsep *Transfer Learning*, model disesuaikan untuk tugas spesifik lewat proses *fine-tuning* [21]. Untuk mencegah terjadinya *overfitting* selama *fine-tuning*, beberapa strategi yang diterapkan meliputi penggunaan *Early Stopping*, *Dropout*, dan *Weight Decay* [22].

### 2.4 Confusion Matrix

*Confusion Matrix* adalah tabel evaluasi yang digunakan untuk mengukur kinerja dari model klasifikasi *machine learning* [23]. Matriks ini memetakan hasil prediksi model ke dalam empat komponen utama:

1. *True Positive* (TP): Komentar spam iklan judi online yang berhasil dideteksi dengan benar.

2. *True Negative* (TN): Komentar normal (*non-spam*) yang dikenali dengan benar.
3. *False Positive* (FP): Komentar normal yang keliru diklasifikasikan sebagai *spam*.
4. *False Negative* (FN): Komentar spam judi online yang gagal dideteksi oleh model.

Berdasarkan keempat nilai tersebut, evaluasi performa model dihitung menggunakan persamaan berikut [24]:

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

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

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

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

### 2.5 Chrome Extension

*Google Chrome Extension* merupakan modul perangkat lunak kecil yang dibangun menggunakan teknologi web standar (HTML, CSS, JavaScript) untuk memodifikasi fungsi peramban [25]. Komponen utamanya, *Content Scripts*, memiliki kemampuan untuk mengakses dan memanipulasi *Document Object Model* (DOM) pada halaman web yang sedang dibuka pengguna [26].

Pada platform dinamis seperti YouTube yang menerapkan sistem *Single Page Application* (SPA), teknologi *MutationObserver* digunakan untuk mendeteksi penambahan elemen komentar baru secara *asynchronous* ketika pengguna melakukan pemindahan halaman atau *scrolling* [27]. Untuk mengatasi keterbatasan komputasi di sisi peramban, pengembangan ekstensi ini mengadopsi arsitektur *Client-Server*, di mana komponen ekstensi bertindak sebagai *frontend* yang mengirimkan data teks via HTTP Fetch API ke server *backend* berbasis Python Flask yang menjalankan model IndoBERT [28].

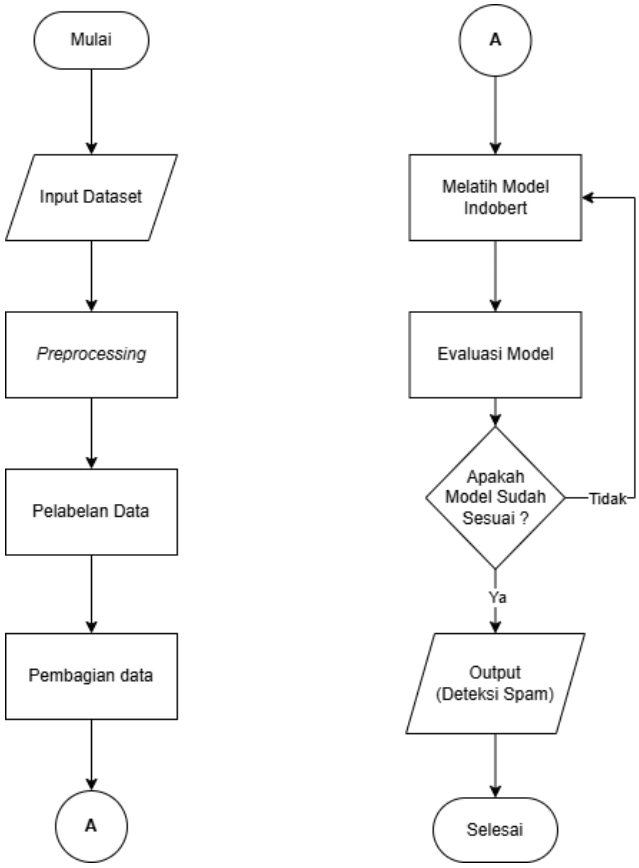
## 3. METODE PENELITIAN

### 3.1 Rancangan Penelitian

Penelitian ini dirancang melalui lima tahapan utama secara sistematis untuk membangun sistem deteksi dan filtrasi komentar spam judi online secara *real-time*. Tahapan penelitian tersebut meliputi:

1. *Data Crawling*: Pengumpulan dataset komentar YouTube berbahasa Indonesia menggunakan YouTube Data API v3.
2. *Preprocessing & Labeling*: Pembersihan teks (*text cleaning*, *case folding*, *normalization*) serta pelabelan semi-terarah (*pseudo-labeling*).
3. *Model Training*: Pelatihan dan *fine-tuning* model IndoBERT serta pengujian *hyperparameter tuning*.
4. *Model Evaluation*: Evaluasi performa model menggunakan *Confusion Matrix*, *Accuracy*, *Precision*, *Recall*, dan *F1-Score*.
5. *System Implementation*: Pengembangan *Chrome Extension* dan integrasi API *backend* berbasis Flask.

Tahapan alur analisis proses penelitian secara keseluruhan ditunjukkan pada Gambar 1.




Gambar 1 Flowchart Analisis Proses

### 3.2 Sampel Penelitian

Sampel data dalam penelitian ini berupa teks komentar video YouTube berbahasa Indonesia yang dikumpulkan melalui dua tahapan *crawling* memanfaatkan YouTube Data API v3 dan library *googleapiclient* pada Python. Pengambilan data difokuskan pada rentang waktu 15 Oktober 2025 hingga 9 Desember 2025 untuk menangkap tren penyamaran teks (*obfuscation*) dan variasi bahasa gaul (*slang*) terbaru.

Total sampel data yang dikumpulkan berjumlah 117.634 komentar yang terbagi menjadi dua kelompok:

1. Dataset Awal (*Seed Dataset*): Sebanyak 3.152 komentar yang diambil secara tertarget menggunakan kata kunci spesifik terkait spam judi (seperti "slot", "gacor", "maxwin", "situs", "judi"). Dataset ini digunakan sebagai basis pelabelan manual.
2. Dataset Utama: Sebanyak 114.482 komentar berskala besar yang dikumpulkan dari 34 video pada *channel* YouTube Deddy Corbuzier untuk kebutuhan strategi *pseudo-labeling*.

### 3.3 Instrumen Penelitian

Instrumen pengumpulan data dalam penelitian ini menggunakan parameter API resmi dan format penyimpanan terstruktur. Komentar yang berhasil ditarik disimpan ke dalam berkas berformat CSV dengan atribut utama berupa teks komentar dan ID video. Untuk menjamin keandalan data sebelum masuk ke tahap pemodelan, instrumen *preprocessing* diterapkan melalui tiga tahapan pembersihan teks:

1. *Text Cleaning*: Menghilangkan elemen non-kontekstual seperti emoji, tautan URL, tanda baca, karakter non-ASCII, serta normalisasi spasi berlebih.
2. *Case Folding*: Menyeragamkan seluruh huruf menjadi huruf kecil (*lowercase*) agar proses analisis kata menjadi konsisten.

3. *Normalization*: Mengubah kata tidak baku atau singkatan media sosial (contoh: "gk", "ga", "skrg", "bngt") menjadi bentuk baku bahasa Indonesia menggunakan kamus substitusi sederhana.

### 3.4 Teknik Pengumpulan dan Analisis Data

Teknik analisis data yang digunakan menggabungkan pendekatan *semi-supervised learning* dua tahap untuk mengatasi masalah ketidakseimbangan data (*imbalanced data*). Proses diawali dengan pelabelan manual terhadap 3.152 data awal menjadi kelas *spam* dan *non-spam*.

Guna memanfaatkan data besar dan menghindari bias, dibangun Model A sebagai pelabel awal menggunakan teknik *Undersampling*, di mana seluruh sampel spam manual diseimbangkan dengan sampel non-spam yang diambil secara acak. Model A kemudian melakukan prediksi otomatis (*pseudo-labeling*) terhadap 114.482 data utama. Hasil akhir dari penggabungan data manual dan otomatis menghasilkan dataset final sebanyak 117.634 data terlabel yang siap digunakan untuk melatih model utama.

### 3.5 Prosedur Pengolahan Data

Prosedur pembagian data dilakukan secara *stratified* (stratifikasi berdasarkan label) untuk menjaga proporsi kelas *spam* dan *non-spam* tetap konsisten di setiap subset. Data gabungan final (117.634 data) dibagi menggunakan tiga skenario rasio eksperimen:

1. Skenario A: 80% Data Latih, 10% Data Validasi, dan 10% Data Uji.
2. Skenario B: 75% Data Latih, 10% Data Validasi, dan 15% Data Uji.
3. Skenario C: 70% Data Latih, 15% Data Validasi, dan 15% Data Uji.

### 3.6 Pembangunan Model IndoBERT

Model utama yang dikembangkan dalam penelitian ini menggunakan arsitektur indobenchmark/indobert-base-p1 yang berbasis *Transformers* dan telah dilatih pada korpus bahasa Indonesia. Proses *fine-tuning* dilakukan pada dataset komentar dengan melakukan adaptasi pada *classification head* untuk klasifikasi dua kelas (spam dan non-spam) dan dioptimasi menggunakan algoritma AdamW. Eksperimen *hyperparameter tuning* diterapkan dengan menguji variasi parameter kunci berikut untuk mencari performa terbaik:

1. *Learning Rate*: [2e-5, 3e-5, 5e-5] (dengan baseline awal 5e-5)
2. *Batch Size*: [16]
3. *Epochs*: [2, 3, 4] (dengan baseline awal 3 epoch)
4. Parameter *weight decay* dijaga konstan sebesar 0,01 memanfaatkan modul Trainer untuk evaluasi otomatis pada setiap epoch.

## 4. HASIL DAN PEMBAHASAN

Spesifikasi lingkungan implementasi sistem melibatkan platform *cloud computing Google Colab* yang didukung oleh akselerasi *hardware* GPU NVIDIA Tesla T4 (16GB VRAM) untuk efisiensi fase pelatihan model final. Adapun implementasi pada sisi pengguna (*client-side*) dijalankan secara lokal pada perangkat berspesifikasi prosesor Intel Core i3-1115G4, RAM 8 GB, dan Windows 11 64-bit melalui peramban *Google Chrome* versi terbaru. Komponen *software* dibangun menggunakan Python 3.10, *framework deep learning* PyTorch dan *Transformers* (Hugging Face), *backend* API Flask, serta *frontend extension* berbasis standar Manifest V3 (HTML, CSS, JavaScript).

Proses pengumpulan data (*crawling*) melalui YouTube Data API v3 berhasil menghimpun total 117.634 komentar mentah. Korpus data tersebut disimpan secara terstruktur ke dalam format Excel (.xlsx). Gambaran umum mengenai variasi struktur teks komentar mentah yang ditarik dari platform YouTube dapat dilihat pada Tabel 1.

2622-8130 Jurnal Sifo Mikroskil (JSM) - Volume 27, No 2, Oktober 2026
DOI: https://doi.org/10.55601/jsm.v27i2.2060

Tabel 1 Contoh Data Mentah Hasil Crawling

| video_id | author | Text |
|---|---|---|
| NIr72bFGFts | @linhliễuhồng1957 | Setiap klik bisa jadi kemenangan—siap buktiin di◉✿🅐🅡🅦🅐🅝🅣🅞🅣🅞🤫💸 |
| NIr72bFGFts | @PUJIYANTOPUJIYANTO-x9h | HOBIQQ<br>tiap hari makin gacor |
| NIr72bFGFts | @EndaLesmana-b4g | "Gacor nya bukan main kalau mainnya di **GELORA4D**" |
| NIr72bFGFts | @goneopinion0.26 | 02:03 Subhanallah, ❤ **BERKAH99** 🔴 bikin takjub! Visualnya kece, ceritanya mengalir penuh makna, dan kemenangannya bikin hati berbunga. Pengalaman yang nggak terlupakan! ✨ |
| NIr72bFGFts | @MultiNurningsih-n1h9f | gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs ϟ TABUNG4D ϟ yang sering gua mainin itu gacor |
| NIr72bFGFts | @bloerykombaet9478 | gacor bertiga klop... bikinin lah mael podcast biar ga rusak badan |
| NIr72bFGFts | @adityastars9186 | Ini kalo masalah agama jangan di kasih musik gokil maellll 5:57 |
| NIr72bFGFts | @iwaydepp | buat bang mael, klo mau tanpa stunman cobain lah bang mael ke tempat gw di Kalimantan buat ngambil madu hutan buat konten |
| NIr72bFGFts | @SeptiAndriyani-q6f | Ya allah aku ketawa sampai keluar air mata🤣🤣🤣🤣🤣 |

Prosedur *preprocessing* data diterapkan secara sistematis melalui tiga tahapan utama untuk menjamin kualitas dan kebersihan data masukan sebelum diproses oleh model IndoBERT. Keberadaan tahapan ini ditunjukkan secara bertahap pada Tabel 2, Tabel 3, dan Tabel 4.

Pertama, tahapan *Text Cleaning* berfokus pada pembersihan elemen non-kontekstual seperti emoji, URL, *user mention*, tanda baca berlebih, serta karakter khusus. Tabel 2 memperlihatkan bagaimana karakter penyamaran (*obfuscation*) yang sering dimanfaatkan oleh *spammer* judi online (seperti penggunaan simbol dekoratif ◉✿, huruf tebal khusus **BERKAH99**, atau emoji) berhasil dihilangkan. Proses ini krusial untuk menyisakan entitas kata utama tanpa adanya gangguan artefak visual.

Tabel 2 Hasil Tahap Text Cleaning

| Teks Asli | Hasil Cleaning |
|---|---|
| Setiap klik bisa jadi kemenangan—siap buktiin di◉✿🅐🅡🅦🅐🅝🅣🅞🅣🅞🤫💸 | Setiap klik bisa jadi kemenangan siap buktiin di |
| HOBIQQ<br>tiap hari makin gacor | QQ tiap hari makin gacor |
| "Gacor nya bukan main kalau mainnya di **GELORA4D**" | Gacor nya bukan main kalau mainnya di |
| 02:03 Subhanallah, ❤ **BERKAH99** 🔴 bikin takjub! Visualnya kece, ceritanya mengalir | Subhanallah bikin takjub Visualnya kece ceritanya mengalir penuh makna dan |

| | |
|---|---|
| penuh makna, dan kemenangannya bikin hati berbunga. Pengalaman yang nggak terlupakan! ✨ | kemenangannya bikin hati berbunga Pengalaman yang nggak terlupakan |
| gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs ⚡ TABUNG4D ⚡ yang sering gua mainin itu gacor | gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs TABUNG D yang sering gua mainin itu gacor |
| gacor bertiga klop... bikinin lah mael podcast biar ga rusak badan | gacor bertiga klop bikinin lah mael podcast biar ga rusak badan |
| Ini kalo masalah agama jangan di kasih musik gokil maellll 5:57 | Ini kalo masalah agama jangan di kasih musik gokil maellll |
| buat bang mael, klo mau tanpa stunman cobain lah bang mael ke tempat gw di Kalimantan buat ngambil madu hutan buat konten | buat bang mael klo mau tanpa stunman cobain lah bang mael ke tempat gw di Kalimantan buat ngambil madu hutan buat konten |
| Ya allah aku ketawa sampai keluar air mata🤣🤣🤣🤣🤣 | Ya allah aku ketawa sampai keluar air mata |

Kedua, tahapan *Case Folding* diterapkan untuk menyeragamkan seluruh huruf dalam teks menjadi huruf kecil (*lowercase*). Tabel 3 menyajikan hasil transformasi teks dari tahap *cleaning* menuju *case folding*. Penyeragaman ini bertujuan agar model IndoBERT memproses kata-kata dengan makna yang sama secara konsisten tanpa terpengaruh oleh variasi kapitalisasi huruf (misalnya mengubah kata "*Subhanallah*" atau "*Gacor*" menjadi huruf kecil seluruhnya).

Tabel 3 Hasil Tahap Case Folding

| Hasil Cleaning | Hasil Case Folding |
|---|---|
| Setiap klik bisa jadi kemenangan siap buktiin di | setiap klik bisa jadi kemenangan siap buktiin di |
| QQ tiap hari makin gacor | qq tiap hari makin gacor |
| Gacor nya bukan main kalau mainnya di | gacor nya bukan main kalau mainnya di |
| Subhanallah bikin takjub Visualnya kece ceritanya mengalir penuh makna dan kemenangannya bikin hati berbunga Pengalaman yang nggak terlupakan | subhanallah bikin takjub visualnya kece ceritanya mengalir penuh makna dan kemenangannya bikin hati berbunga pengalaman yang nggak terlupakan |
| gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs TABUNG D yang sering gua mainin itu gacor | gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs tabung d yang sering gua mainin itu gacor |
| gacor bertiga klop bikinin lah mael podcast biar ga rusak badan | gacor bertiga klop bikinin lah mael podcast biar ga rusak badan |
| Ini kalo masalah agama jangan di kasih musik gokil maellll | ini kalo masalah agama jangan di kasih musik gokil maellll |
| buat bang mael klo mau tanpa stunman cobain lah bang mael ke tempat gw di Kalimantan buat ngambil madu hutan buat konten | buat bang mael klo mau tanpa stunman cobain lah bang mael ke tempat gw di kalimantan buat ngambil madu hutan buat konten |
| Ya allah aku ketawa sampai keluar air mata | ya allah aku ketawa sampai keluar air mata |

Ketiga, tahapan *Normalization* berfokus pada pembenahan kosakata tidak baku, singkatan khas media sosial, maupun ejaan gaul bahasa Indonesia (seperti mengubah singkatan "*bgt*" menjadi "*banget*", "*ga*" menjadi "*tidak*", atau "*klo*" menjadi "*kalau*"). Hasil pemetaan standarisasi kata ini disajikan secara lengkap pada Tabel 4. Sementara itu, istilah asing atau kata berbahasa Inggris yang sering muncul dalam promosi judi online (seperti *jackpot*, *link*, *free*, atau *bonus*) sengaja dipertahankan karena merupakan fitur pembeda semantik yang sangat kuat untuk membantu model mengenali pola kalimat spam judi.

Tabel 4 Hasil Tahap Normalization

| Hasil Case Folding | Hasil Normalization |
|---|---|
| setiap klik bisa jadi kemenangan siap buktiin di | setiap klik bisa jadi kemenangan siap buktiin di |
| qq tiap hari makin gacor | qq tiap hari makin gacor |
| | |
| gak sabar share ke temen soal ceritanya bikin penasaran kemenangannya bikin happy | tidak sabar share ke temen soal ceritanya membuat penasaran kemenangannya membuat happy |
| subhanallah bikin takjub visualnya kece ceritanya mengalir penuh makna dan kemenangannya bikin hati berbunga pengalaman yang nggak terlupakan | subhanallah membuat takjub visualnya kece ceritanya mengalir penuh makna dan kemenangannya membuat hati berbunga pengalaman yang nggak terlupakan |
| gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share situs tabung d yang sering gua mainin itu gacor | gamau banyak bacot gua bang nanti di kira sales wkwk gua cuman mau share website tabung d yang sering gua mainin itu gacor |
| gacor bertiga klop bikinin lah mael podcast biar ga rusak badan | gacor bertiga klop bikinin lah mael podcast biar tidak rusak badan |
| ini kalo masalah agama jangan di kasih musik gokil maellll | ini kalau masalah agama jangan di kasih musik gokil maellll |
| buat bang mael klo mau tanpa stunman cobain lah bang mael ke tempat gw di kalimantan buat ngambil madu hutan buat konten | buat bang mael klo mau tanpa stunman cobain lah bang mael ke tempat saya di kalimantan buat ngambil madu hutan buat konten |
| ya allah aku ketawa sampai keluar air mata | ya allah saya ketawa sampai keluar air mata |

Guna mengatasi isu ketidakseimbangan data (*imbalanced data*) dari hasil pelabelan manual dan otomatis (*pseudo-labeling*), diterapkan teknik *undersampling* pada kelas mayoritas (*Non-Spam*) dengan proporsi batas maksimal rasio 1:3. Langkah ini menjaga variasi kebahasaan normal agar model tetap mampu mengenali konteks percakapan asli tanpa terdominasi oleh kelas mayoritas. Dataset final berukuran 68.472 sampel data yang terbagi atas 17.118 data Spam (Label 1) dan 51.354 data Non-Spam (Label 0).

Pengujian skenario pembagian data menunjukkan bahwa Skenario B dengan rasio proporsi 75% Data Latih, 10% Data Validasi, dan 15% Data Uji memberikan hasil evaluasi terbaik. Hasil pengujian ini mencatatkan nilai akurasi sebesar 93,42% dan *F1-Score* sebesar 86,69%. Alokasi data uji sebesar 15% terbukti memadai dan representatif dalam mengukur kemampuan generalisasi model dibandingkan Skenario A (rasio data uji 10%) yang rentan memicu indikasi *overfitting* ringan, ataupun Skenario C (rasio data latih 70%) yang menurunkan sensitivitas model dalam merekam variasi pola kalimat spam yang kompleks. Rincian perbandingan antar skenario disajikan pada Tabel 5.

Tabel 5 Perbandingan Hasil Skenario Pembagian Data

| Skenario | Rasio | Akurasi (%) | F1-Score (%) |
|---|---|---|---|
| Skenario A | 80:10:10 | 92.71 | 84.76 |
| Skenario B | 75:10:15 | 93.42 | 86.69 |
| Skenario C | 70:15:15 | 93.34 | 86.36 |

Tahap eksperimen *hyperparameter tuning* pada sampel data representatif (10.271 data) menetapkan konfigurasi Eksperimen 1 (Exp-1) sebagai parameter optimal, yaitu *Learning Rate* sebesar 2e-5, jumlah 3 *Epoch*, dan *Batch Size* konstan 16. Nilai *learning rate* yang kecil terbukti mencegah fenomena *overshooting* (melewati titik konvergensi optimal). Penambahan jumlah iterasi hingga *Epoch* 4 justru memicu penurunan nilai *F1-Score* dari 87,84% menjadi 85,30% yang mengindikasikan awal terjadinya gejala *overfitting*. Penggunaan *Batch Size* 16 dipilih sebagai batas aman alokasi memori

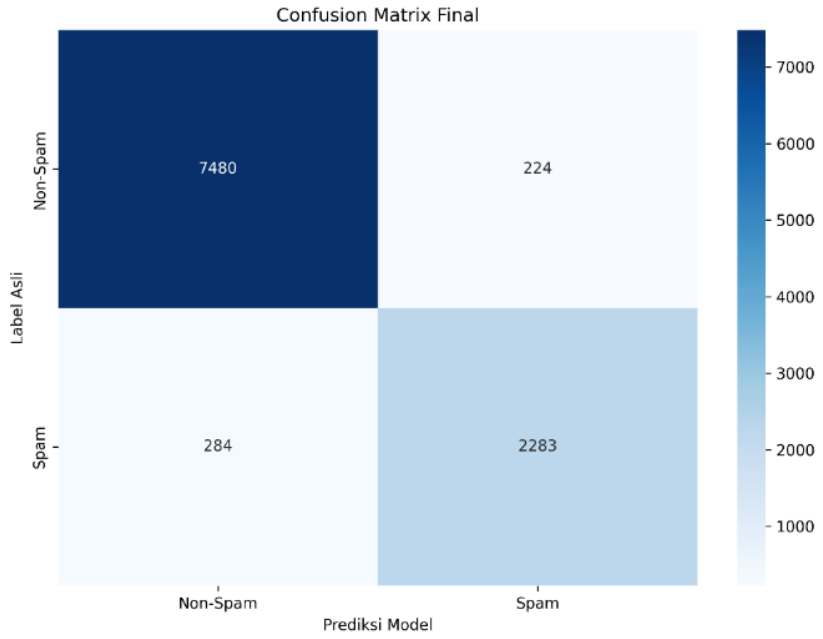
VRAM GPU untuk mengantisipasi kegagalan sistem (*Out of Memory*). Hasil eksperimen *hyperparameter tuning* diperlihatkan secara rinci pada Tabel 6.

Tabel 6 Hasil Eksperimen Hyperparameter Tuning

| Eksperimen | Learning Rate | Epoch | Batch Size | Akurasi | F1-Score |
|---|---|---|---|---|---|
| Exp-1 | 0,00002 | 3 | 16 | 0,939649578 | 0,878431373 |
| Exp-2 | 0,00003 | 2 | 16 | 0,936404932 | 0,874680307 |
| Exp-3 | 0,00002 | 2 | 16 | 0,933809215 | 0,869230769 |
| Exp-4 | 0,00003 | 3 | 16 | 0,92926671 | 0,854860186 |
| Exp-5 | 0,00002 | 4 | 16 | 0,927319922 | 0,853018373 |
| Exp-6 | 0,00003 | 4 | 16 | 0,923426347 | 0,845953003 |
| Exp-7 | 0,00005 | 2 | 16 | 0,919532771 | 0,837270341 |
| Exp-8 | 0,00005 | 3 | 16 | 0,912394549 | 0,82260184 |
| Exp-9 | 0,00005 | 4 | 16 | 0,909798832 | 0,813422819 |

Pelatihan model final (*fine-tuning*) menggunakan arsitektur *pre-trained* indobenchmark/indobert-base-p1 pada dataset penuh menunjukkan peningkatan performa semantik yang konsisten. Nilai *F1-Score* bergerak naik beriringan dengan nilai akurasi hingga mencapai titik optimalnya sebesar 90,1% pada *Epoch* 3. Bobot model pada titik ini disimpan sebagai model final untuk digunakan pada pengujian data uji (*testing set*).

Evaluasi performa model final menggunakan 10.271 data uji mandiri menghasilkan nilai akurasi global sebesar 95,05% dengan capaian metrik kelas Spam berupa Presisi 91%, *Recall* 89%, dan *F1-Score* 90%. Berdasarkan pemetaan *Confusion Matrix* pada Gambar 2, sistem mencatatkan 7.480 *True Negative* (TN), 2.283 *True Positive* (TP), 224 *False Positive* (FP/Kesalahan Tipe I), dan 284 *False Negative* (FN/Kesalahan Tipe II). Nilai presisi yang tinggi (91%) menjamin tingkat *false alarm* yang rendah sehingga aman meminimalkan risiko salah blokir terhadap komentar asli pengguna, sedangkan nilai *Weighted Average* yang setara dengan akurasi global menegaskan stabilitas performa model pada data tidak seimbang. Rincian metrik per kelas disajikan dalam *Classification Report* pada Tabel 7.




Gambar 2 Confusion Matrix Model IndoBERT pada Data Uji

Tabel 7 Classification Report Model IndoBERT

| Kelas | Precision | Recall | F1-Score | Jumlah Data (Support) |
|---|---|---|---|---|
| Non-Spam | 96% | 97% | 97% | 7.704 |
| Spam | 91% | 89% | 90% | 2.567 |
| Akurasi | | | 95% | 10.271 |
| Macro Avg | 94% | 93% | 93% | 10.271 |
| Weighted Avg | 95% | 95% | 95% | 10.271 |

Hasil analisis kualitatif terhadap kesalahan klasifikasi (*error analysis*) memetakan kelemahan model ke dalam dua aspek utama. Kesalahan *False Positive* dipicu oleh bias kontekstual model terhadap kata "*cuan*" (aspek finansial), kosakata doa seperti "*alhamdulilah*" dan "*berkah*" (karena *spammer* kerap memanipulasi kedok *giveaway* syariah), serta kata *slang* "*auto*" yang diidentifikasi kaku sebagai fitur spam judi online. Rincian sampel *False Positive* disajikan pada Tabel 8.

Tabel 8 Sampel Analisis False Positive

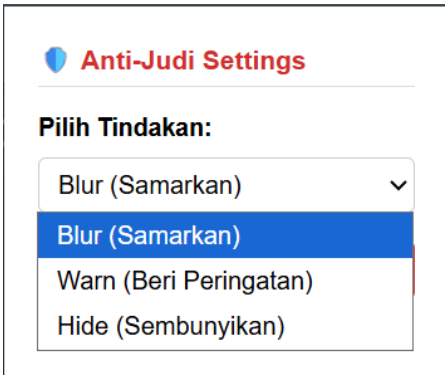
| No | Teks Komentar (Asli) | Analisis Penyebab Error |
|---|---|---|
| 1 | *"indonesia ladang cari cuan habib group"* | Kata "cuan" (untung) sangat dominan dalam data latih spam judi online. Model menjadi sensitif dan menganggap kalimat yang membahas keuntungan finansial sebagai indikator spam. |
| 2 | *"alhamdulilah ya allah semoga berkah dunia akhirat aamiin"* | Ini adalah contoh klasik bias model terhadap kata-kata doa ("alhamdulilah", "berkah", "aamiin"). Karena banyak spammer menggunakan kedok "sedekah/giveaway syariah", model kesulitan membedakan antara doa tulus dari pengguna asli dengan pancingan penipuan. |
| 3 | *"buat tongkrongan tengah malem auto mpe subuh ni"* | Bias Kata Slang: Kata "auto" sering muncul dalam spam judi (*"auto win"*, *"auto gacor"*). Di sini, pengguna menggunakannya sebagai bahasa gaul (*slang*) untuk "otomatis/pasti". Model gagal menangkap konteks gaul ini dan melabelinya sebagai fitur spam. |

Sebaliknya, kesalahan *False Negative* dipicu oleh teknik *social engineering* dan promosi terselubung (*soft-selling*), manipulasi gaya bahasa berupa testimoni curhatan personal pengguna, serta struktur frasa yang ambigu seperti penggunaan istilah "*tanggal main*". Rincian sampel *False Negative* disajikan pada Tabel 9.

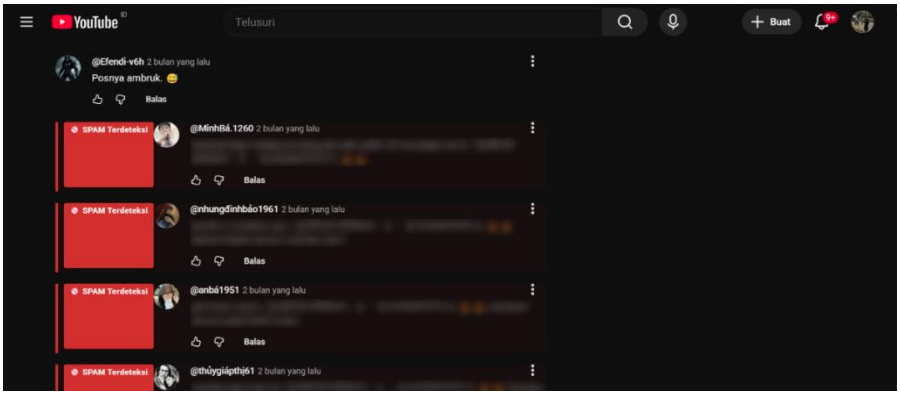
Tabel 9 Sampel Analisis False Negative

| No | Teks Komentar (Asli) | Analisis Penyebab Error |
|---|---|---|
| 1 | *"semua berkat sponsor nya jadi seru"* | Komentar ini menggunakan teknik *soft-selling.* Spammer hanya menyebut "sponsor" secara halus untuk memancing rasa penasaran. Model yang dilatih untuk mendeteksi spam agresif gagal mengenali pola promosi halus ini. |
| 2 | *"saya gapercaya rezeki datang cepat tapi di semua bisa terjdi"* | Kalimat ini disusun menyerupai testimoni atau curhatan pengalaman pribadi, yang merupakan teknik umum dalam promosi judi online. Karena struktur kalimatnya mirip dengan narasi pengguna biasa dan minim kata kunci "blacklist", model mengklasifikasikannya sebagai *Non-Spam*. |
| 3 | *"ditunggu tanggal main nya sudah pokok nunggu pake bgt"* | Frasa “tanggal main” bersifat ambigu. Dalam konteks normal bisa berarti jadwal acara, namun dalam konteks spam sering merujuk pada ”waktu menang” perjudian. Tanpa konteks eksternal (seperti tautan profil pengirim), model menganggap komentar ini sebagai antusiasme penggemar biasa. |

Sistem deteksi akhir diimplementasikan secara nyata ke dalam *Chrome Extension* menggunakan arsitektur *Client-Server* melalui integrasi API backend Flask dan komponen *Content Script* berbasis JavaScript. Saat pengguna membuka halaman YouTube, *Content Script* secara dinamis memantau munculnya komentar baru melalui fungsi *MutationObserver*. Sistem secara *real-time* mengambil preferensi pengguna dari menu popup dan mengirimkan request data ke server lokal untuk dievaluasi oleh model IndoBERT. Manipulasi struktur visual DOM (*Document Object Model*) halaman YouTube dilakukan secara instan tanpa proses memuat ulang halaman (*refresh*) melalui mekanisme *message passing*. Antarmuka menyediakan fleksibilitas penuh bagi pengguna melalui tiga opsi tindakan mitigasi otomatis: menyamarkan teks komentar (*Blur*), menghilangkan elemen teks dari halaman (*Hide*), atau menyematkan penanda visual bingkai merah (*Warn*). *Interface* menu pengaturan *popup* ditunjukkan pada Gambar 3, sedangkan hasil tindakan mitigasi *real-time* dengan mode *Blur* pada halaman YouTube diperlihatkan pada Gambar 4.




Gambar 3 Tampilan Menu Pengaturan (Popup) pada Chrome Extension




Gambar 4 Implementasi Deteksi Real-time dengan Mode Blur

Pengujian fungsionalitas sistem secara menyeluruh (*End-to-End*) diselesaikan melalui dua tahapan pengujian. Pengujian tahap pertama difokuskan pada aspek antarmuka pengguna (*user interface*) menggunakan metode *Black Box Testing*. Melalui skenario pengujian langsung (*live testing*) pada kolom komentar saluran YouTube Raditya Dika, ekstensi terbukti andal mengeksekusi fungsi mitigasi visual (mode *Blur*, interaksi *Hover* untuk mengintip teks asli, dan mode *Hide*) pada lingkungan riil di luar korpus data pelatihan.

Pengujian tahap kedua difokuskan pada stabilitas integrasi komponen API backend melalui pengujian otomatis (*Automated Testing*) memanfaatkan aplikasi Postman. Hasil pengujian menunjukkan bahwa server backend sukses merespons permintaan JSON dengan status kode HTTP 200 OK serta mengembalikan nilai probabilitas prediksi yang akurat berupa array label {"predictions": [1]}. Capaian indikator kelulusan penuh (3/3 *Passed*) pada lembar *Test Results* menegaskan bahwa jembatan

komunikasi data antara ekstensi peramban di sisi klien dan model deep learning IndoBERT di sisi server telah terintegrasi dengan sempurna.

## 5. KESIMPULAN

Penelitian ini berhasil mengimplementasikan aplikasi deteksi komentar spam judi online pada platform YouTube berbahasa Indonesia menggunakan model IndoBERT dan *Chrome Extension*. Model IndoBERT yang di-*fine-tune* terbukti andal mengatasi ketidakseimbangan kelas (*imbalanced dataset*) dengan capaian akurasi sebesar 95,05%, F1-Score 90%, dan presisi tinggi sebesar 91% yang meminimalkan risiko salah tangkap (*False Positive*). Melalui *preprocessing* yang tepat, model efektif mengenali variasi bahasa gaul (*slang*), singkatan, dan teknik pengelabuan kata (*obfuscation*), meskipun masih memiliki sedikit bias terhadap kosakata religius dan finansial. Aplikasi *Chrome Extension* berbasis Manifest V3 dengan arsitektur *Client-Server* sukses mengekstraksi dan memanipulasi DOM YouTube secara *real-time* lewat fitur *Blur, Hide,* atau *Warn* tanpa perlu memuat ulang halaman. Sistem ini menghadirkan solusi praktis untuk memitigasi sebaran konten judi online demi menciptakan ekosistem digital yang aman.

## 6. SARAN

Berdasarkan hasil evaluasi kinerja sistem dan identifikasi keterbatasan selama proses eksperimen berlangsung, penelitian selanjutnya disarankan untuk mempertimbangkan beberapa aspek pengembangan berikut:

1. Infrastruktur & Data: Memanfaatkan GPU berspesifikasi lebih tinggi untuk mengeksplorasi *batch size* yang lebih besar, serta memperkaya variasi dataset pada komentar non-spam yang memuat kata sensitif (religius dan finansial) guna mengurangi bias model.
2. Arsitektur Sistem: Mentransformasikan sistem ke arah *Edge Computing* dengan mengonversi model ke format ONNX/TensorFlow.js atau menggunakan DistilBERT agar proses inferensi berjalan langsung di sisi *browser* untuk memangkas latensi jaringan.
3. Fitur Ekstensi: Mengembangkan model *Dual-Input* untuk menganalisis teks komentar sekaligus nama akun (*username*), serta mengintegrasikan *feedback loop* berbasis *Active Learning* agar pengguna dapat melaporkan spam baru secara manual sebagai data pelatihan ulang (*re-training*) model secara berkala.

## DAFTAR PUSTAKA

[1] R. Tjahyadi, V. C. Mawardi, and N. J. Perdana, “Aplikasi Deteksi Komentar Spam Youtube Dengan Metode Support Vector Machine Berbasis Web,” *Jurnal Serina Sains, Teknik dan Kedokteran*, vol. 1, no. 2, pp. 237–246, Sep. 2023, doi: 10.24912/jsstk.v1i2.28510.

[2] Arif Nur Cahyo, Riky Hermawan, Moh Avin Dharma Wijaya, and Anggraini Puspita Sari, “Deteksi Teks Promosi Judi Online Pada Kolom Komentar YouTube dengan Metode Regex dan Fuzzy Matching,” *Jurnal RESTIKOM : Riset Teknik Informatika dan Komputer*, vol. 7, no. 2, pp. 176–187, 2025, doi: 10.52005/restikom.v7i2.441.

[3] A. D. L. Uro Abdurohim, Dedy Apriyadi, “Implementasi Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Komentar Spam Pada Instagram,” vol. 13, no. 1, pp. 13–19, 2024.

[4] D. P. K. Samuel, “Deteksi Teks Promosi Judi Online Menggunakan Ai Dengan Kombinasi Nlp Dan Deep Learning,” vol. V, no. 2, pp. 179–185.

[5] K. Kamdan, M. P. Anugrah, M. J. Almutaali, R. Ramdani, and I. L. Kharisma, “Performance Analysis of IndoBERT for Detection of Online Gambling Promotion in YouTube Comments,” MDPI AG, Sep. 2025, p. 66. doi: 10.3390/engproc2025107066.

[6] F. Jauhari, M. Riza, R. G. Guntara, and M. R. Nugraha, "Implementasi Algoritma Naive Bayes untuk Filtrasi Spam Komentar Judi Online pada YouTube," *Indonesian Journal of Digital Business*, vol. 5, no. 2, pp. 411–423, 2025, [Online]. Available: https://ejournal.upi.edu/index.php/IJDB/article/view/88641

[7] N. Fikriah Nst, D. Hamdhana, and M. Qamal, "Few-Shot Learning for Classifying Genuine and Bot Comments on YouTube Using Transformer Models," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1578–1586, 2025, doi: 10.30871/jaic.v9i4.10023.

[8] A. Widiyanto, M. Prameswari, and M. Abdul Latief, "Gambling Comments Detection on Youtube: A Comparative Study of Tree-Based Boosting, LSTM and GRU Models," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 23, no. 2, pp. 144–160, 2025, doi: 10.12962/j24068535.v23i2.a1305.

[9] H. Alkaabi, A. K. Jasim, and A. Darroudi, "From Static to Contextual: A Survey of Embedding Advances in NLP," *PERFECT: Journal of Smart Algorithms*, vol. 2, no. 2, pp. 57–66, 2025, doi: 10.62671/perfect.v2i2.77.

[10] F. Rifaldy, Y. Sibaroni, and S. S. Prasetiyowati, "Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 863–874, 2025, doi: 10.29100/jipi.v10i2.6055.

[11] U. N. Jakarta, V. B. Lestari, and C. A. Hutagalung, "J-KOMA Journal of Computer Science and Applications Evaluation of TF-IDF Extraction Techniques in Sentiment Analysis of Indonesian-Language Marketplaces Using SVM, Logistic Regression, and Naive Bayes," pp. 22–2025, doi: 10.21009/j.

[12] Sutriawan, S. Rustad, G. F. Shidik, and Pujiono, "Performance Evaluation of Text Embedding Models for Ambiguity Classification in Indonesian News Corpus: A Comparative Study of TF-IDF, Word2Vec, FastText BERT, and GPT," *Ingenierie des Systemes d'Information*, vol. 30, no. 6, pp. 1469–1482, 2025, doi: 10.18280/isi.300606.

[13] Ibnu Daqiqil Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python.*, 1st ed. Gobah Pekanbaru, 2021.

[14] M. Sri, L. Sajja, L. Yuktha, N. Esther, and P. Charitha, "Enhancing Spam Comment Detection On Social Media With Emoji Feature And Post-Comment Pairs Approach Using Ensemble Methods Of Machine Learning", doi: 10.62646/ijitce.2024.v15.i3.pp837-845.

[15] G. Airlangga, "Spam Detection in YouTube Comments Using Deep Learning Models: A Comparative Study of MLP, CNN, LSTM, BiLSTM, GRU, and Attention Mechanisms," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1533–1538, Oct. 2024, doi: 10.57152/malcom.v4i4.1671.

[16] K. Kadarsih and D. Pujianto, "Application Of Transformer Model And Word Embedding In Sentiment Analisys Of Indonesian E-Commerce Application Review," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 7, no. 3, pp. 720–732, Jul. 2025, doi: 10.47709/cnahpc.v7i3.6354.

[17] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, pp. 757–770, 2020, doi: 10.18653/v1/2020.coling-main.66.

[18] S. M. Isa, G. Nico, and M. Permana, "Indobert for Indonesian Fake News Detection," *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, 2022, doi: 10.24507/icicel.16.03.289.

[19] Y. Liang and M. Li, "A deep learning model for prediction of lysine crotonylation sites by fusing multi-features based on multi-head self-attention mechanism," *Sci. Rep.*, vol. 15, no. 1, pp. 1–12, 2025, doi: 10.1038/s41598-025-04058-5.

[20] W. Liang and Y. Liang, "BPDec: Unveiling the Potential of Masked Language Modeling Decoder in BERT Model Pretraining," *Lecture Notes in Computer Science*, vol. 15294 LNCS, no. Mlm, pp. 393–408, 2025, doi: 10.1007/978-981-96-6599-0_27.

[21] C. Shaw, P. LaCasse, and L. Champagne, "Exploring emotion classification of indonesian tweets using large scale transfer learning via IndoBERT," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, Dec. 2025, doi: 10.1007/s13278-025-01439-6.

[22] I. Markoulidakis and G. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies (Basel).*, vol. 12, no. 7, Jul. 2024, doi: 10.3390/technologies12070113.

[23] A. Andryani, A. N. Salim, T. Sutabri, and K. Kunci, "Deteksi Email Spam dan Non-Spam Berdasarkan Isi Konten Menggunakan Metode K-Nearest Neighbor dan Support Vector Machine," vol. 6, no. 2, 2024, doi: 10.46799/syntax.

[24] M. Sitarz, "Extending F1 Metric, Probabilistic Approach," *Advances in Artificial Intelligence and Machine Learning*, vol. 3, no. 2, pp. 1025–1038, 2023, doi: 10.54364/AAIML.2023.1161.

[25] G. Girinoto, D. A. Alwan, G. A. N. Gde K.T. D, O. G. Nabila, A. Arizal, and D. F. Priambodo, "Implementasi Deteksi Judul Berita Clickbait Berbahasa Indonesia dengan pre-trained model Multilingual BERT Pada Aplikasi Berbasis Chrome Extension," *Jurnal Ilmiah SINUS*, vol. 20, no. 2, p. 25, Jul. 2022, doi: 10.30646/sinus.v20i2.624.

[26] M. Aljabri *et al.*, "Kashif: A Chrome Extension for Classifying Arabic Content on Web Pages Using Machine Learning," *Applied Sciences (Switzerland)*, vol. 14, no. 20, Oct. 2024, doi: 10.3390/app14209222.

[27] S. Essahli, O. Sarsar, I. Fouad, A. Bentajer, and A. Motii, "FakeZero: Real-Time, Privacy-Preserving Misinformation Detection for Facebook and X," *Proceedings of the IEEE International Conference on Trust, Security and Privacy in Computing and Communications, TrustCom*, no. 2025, pp. 2025–2035, 2025, doi: 10.1109/Trustcom66490.2025.00235.

[28] M. Raheem, N. F. Abubacker, and D. Wiyani, "Sentiment-Based Recommendation for Online Shopping," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 9, pp. 4196–4207, 2024.

2622-8130
Jurnal Sifo Mikroskil (JSM) - Volume 27, No 2, Oktober 2026
DOI: https://doi.org/10.55601/jsm.v27i2.2060