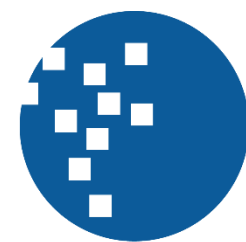


**ANALISIS PENERAPAN INDOBERT DAN FUZZY MATCHING
UNTUK SISTEM DETEKSI TEKS JUDI ONLINE
PADA MEDIA SOSIAL**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tugas Akhir

Ian Pangeswara Hermawan

00000055622

**PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**ANALISIS PENERAPAN INDOBERT DAN FUZZY MATCHING
UNTUK SISTEM DETEKSI TEKS JUDI ONLINE
PADA MEDIA SOSIAL**



Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Teknik

Ian Pangeswara Hermawan

00000055622

**PROGRAM STUDI TEKNIK KOMPUTER
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Ian Pangeswara Hermawan

Nomor Induk Mahasiswa : 00000055622

Program Studi : Teknik Komputer

Menyatakan dengan sesungguhnya bahwa Tugas Akhir saya yang berjudul:

ANALISIS PENERAPAN INDOBERT DAN FUZZY MATCHING UNTUK
SISTEM DETEKSI TEKS JUDI ONLINE PADA MEDIA SOSIAL

Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 10 Juni 2025

A pink rectangular stamp with a QR code and the text "METERAI ELEKTRONIK 10000 RUPIAH" is visible. Overlaid on the stamp is a handwritten signature in black ink.

Ian Pangeswara Hermawan

HALAMAN PENGESAHAN

Tugas Akhir dengan judul

ANALISIS PENERAPAN INDOBERT DAN FUZZY MATCHING UNTUK
SISTEM DETEKSI TEKS JUDI ONLINE PADA MEDIA SOSIAL

Oleh

Nama : Ian Pangeswara Hermawan
NIM : 00000055622
Program Studi : Teknik Komputer
Fakultas : Teknik dan Informatika

Telah diujikan pada hari Selasa, 10 Juni 2025
Pukul 10.00 s.d 12.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut.

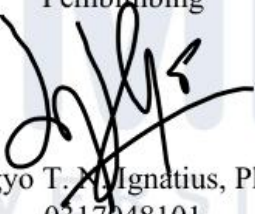
Ketua Sidang


Monica Pratiwi, S.ST., M.T.
0325059601

Penguji


Samuel Hutagalung, M.T.I.
0304038902

Pembimbing


Hargyo T. M. Ignatius, Ph.D.
0317048101

Ketua Program Studi Teknik Komputer


Samuel Hutagalung, M.T.I.
0304038902

iii

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Ian Pangeswara Hermawan
NIM : 00000055622
Program Studi : Teknik Komputer
Jenjang : S1
Judul Karya Ilmiah : Analisis Penerapan IndoBERT Dan Fuzzy
Matching Untuk Sistem Deteksi Teks Judi
Online Pada Media Sosial

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - ☐ Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - ☐ Embargo publikasi karya ilmiah dalam kurun waktu 3 tahun.

Tangerang, 28 Mei 2025



Ian Pangeswara Hermawan

** Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa atas karuniaNya sehingga penulisan laporan tugas akhir dapat diselesaikan dengan baik. Dengan ini penulis ingin menyampaikan terima kasih kepada berbagai pihak yang membantu antara lain:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Ibu Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
3. Bapak Samuel Hutagalung, M.T.I., selaku Ketua Program Studi Teknik Komputer, Universitas Multimedia Nusantara.
4. Bapak Hargyo T. N. Ignatius, Ph.D., selaku Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi sehingga terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman saya Bayu Baskoro Adjie, Rafi Rabbani, Erland Hari Nofiar, Natasya Syafa Salsabilla, Roihan Fadhlurrohman, Nur Sandi, Evan Yulius Simanjuntak, dan Hendie Jason Achmadi yang telah mendukung dan memotivasi dalam pengerjaan laporan tugas akhir baik secara langsung maupun tidak.

Semoga laporan tugas akhir ini dapat bermanfaat, baik sebagai sumber informasi, dan menjadi rujukan bagi penelitian berikutnya.

Tangerang, 28 Mei 2025



Ian Pangeswara Hermawan

ANALISIS PENERAPAN INDOBERT DAN FUZZY MATCHING UNTUK SISTEM DETEKSI TEKS JUDI ONLINE PADA MEDIA SOSIAL

(Ian Pangeswara Hermawan)

ABSTRAK

Kasus perjudian di masyarakat telah menjadi permasalahan yang tidak pernah kunjung selesai sejak bertahun-tahun lamanya, yang sekarang berkembang menjadi judi online. Faktor utama penyebab maraknya perjudian online adalah teknologi dan kemudahan akses, terutama media sosial. Data menunjukkan 82% orang yang mengakses internet pernah melihat iklan judi online. Salah satu upaya untuk mengurangi iklan judi online di media sosial adalah dengan penggunaan teknologi machine learning dan artificial intelligence (AI) untuk mendeteksi teks yang terindikasi judi online. Beberapa penelitian terdahulu membuktikan metode seperti IndoBERT dan Fuzzy Matching cocok digunakan dalam mengklasifikasi dan mengidentifikasi sebuah teks. Metode Fuzzy Matching dapat digunakan sebagai mekanisme awal untuk pelabelan dataset berbasis kemiripan kata, sementara IndoBERT dilatih untuk memahami konteks kalimat. Tiga skenario diuji: IndoBERT murni, Fuzzy Matching murni, dan model hybrid untuk melihat pada kondisi seperti apa model memiliki performa baik. Hasil menunjukkan bahwa pendekatan hybrid menghasilkan performa paling stabil dengan F1-score 0,77 pada data teks tidak baku dan akurasi di atas 90% pada data bersih. Meski demikian, performa menurun pada data yang benar-benar baru, mengindikasikan keterbatasan generalisasi akibat penggunaan kata kunci yang bersifat statis.

Kata kunci: Judi Online, Media Sosial, IndoBERT, *Fuzzy Matching*

ANALYSIS OF THE IMPLEMENTATION OF INDOBERT AND FUZZY MATCHING FOR ONLINE GAMBLING TEXT DETECTION SYSTEM ON SOCIAL MEDIA

(Ian Pangeswara Hermawan)

ABSTRACT (English)

Gambling cases in society have been a problem that has never been resolved for years, which has now developed into online gambling. The main factors causing the rise of online gambling are technology and ease of access, especially social media. Data shows that 82% of people who access the internet have seen online gambling advertisements. One effort to reduce online gambling advertisements on social media is to use machine learning and artificial intelligence (AI) technology to detect texts that indicate online gambling. Several previous studies have shown that methods such as IndoBERT and Fuzzy Matching are suitable for classifying and identifying text. The Fuzzy Matching method can be used as an initial mechanism for labeling datasets based on word similarity, while IndoBERT is trained to understand the context of sentences. Three scenarios were tested: pure IndoBERT, pure Fuzzy Matching, and a hybrid model to see under what conditions the model performs well. The results show that the hybrid approach produces the most stable performance with an F1-score of 0.77 on non-standard text data and accuracy above 90% on clean data. However, performance decreases on completely new data, indicating limited generalization due to the use of static keywords.

Keywords: Online Gambling, Social Media, IndoBERT, Fuzzy Matching

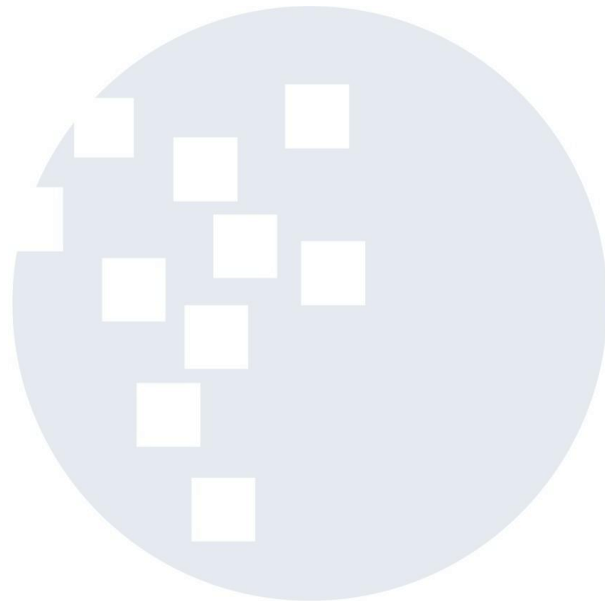
DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
ABSTRAK	vi
<i>ABSTRACT (English)</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	3
1.3 Batasan Penelitian	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.1.1 Penelitian Terkait Judi Online	6
2.1.2 Penelitian Terkait BERT	8
2.1.3 Penelitian Terkait Fuzzy Matching	9
2.2 Tinjauan Teori	11
2.2.1 <i>Text Classification</i>	11
2.2.2 <i>NLP</i>	11
2.2.3 <i>Fuzzy Matching</i>	11
2.2.4 <i>BERT</i>	12
BAB III ANALISIS DAN PERANCANGAN SISTEM	13
3.1 Metode Penelitian	13
3.2 Studi Literatur	13

3.3 Koleksi & Pembagian Dataset	13
3.3.1 Dataset Tipe 1	14
3.3.2 Dataset Tipe 2	14
3.3.3 Dataset Tipe 3	15
3.3.4 Dataset Training	16
3.4 Perancangan Skenario Pengujian	16
3.5 Tahapan Pengujian Pada Skenario 1 – 9	18
3.5.1 Skenario 1 – 3	18
3.5.2 Skenario 4 - 6	20
3.5.3 Skenario 7 - 9	21
3.6 Metriks Evaluasi	22
BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM	23
4.1 Spesifikasi Sistem	23
4.2 Implementasi Sistem	23
4.2.1 Import & Download Libraries	23
4.2.2 Pre-Processing Dataset Training	24
4.2.3 Skenario 1 - 3	28
4.2.4 Skenario 4 – 6	32
4.2.5 Skenario 7 - 9	33
4.3 Analisis & Hasil Pengujian Sistem	34
4.3.1 Analisis & Hasil Skenario 1 - 3	34
4.3.2 Analisis & Hasil Skenario 4 – 6	37
4.3.3 Analisis & Hasil Skenario 7 – 9	40
4.3.4 Evaluasi Hasil Skenario 7 - 9	42
BAB V SIMPULAN DAN SARAN	49
5.1 Simpulan	49
5.2 Saran	49
DAFTAR PUSTAKA	51

DAFTAR TABEL

Tabel 3.1 Perbedaan Metode, Labelling, dan Dataset Skenario 1 – 9.....	16
--	----



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

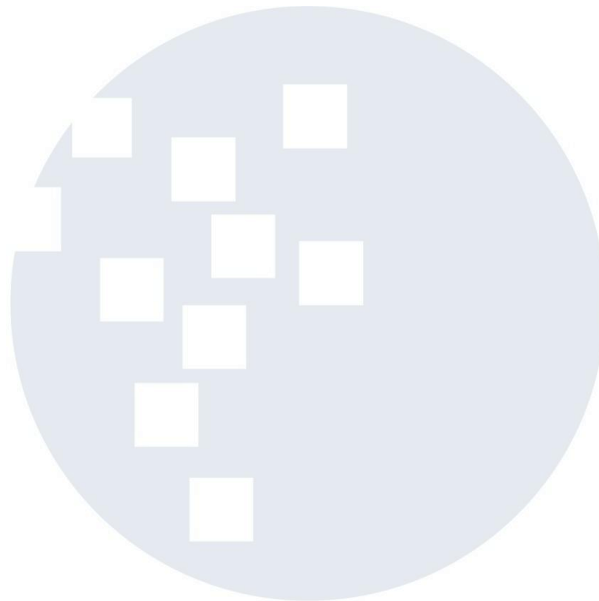
DAFTAR GAMBAR

Gambar 3.1 Flowchart metode penelitian.....	13
Gambar 3.2 Gambaran dataset tipe 1.....	14
Gambar 3.3 Gambaran dataset tipe 2.....	15
Gambar 3.4 Gambaran dataset tipe 3.....	15
Gambar 3.5 Dataset Training.....	16
Gambar 3.6 Flowchart Skenario 1 – 3.....	18
Gambar 3.7 Kolom cleaned_message yang dibutuhkan.....	18
Gambar 3.8 Flowchart Skenario 4 – 6.....	20
Gambar 3.9 Flowchart Skenario 7 – 9.....	21
Gambar 4.1 Import & Download Library Gambar.....	24
Gambar 4.2 Hasil Sebelum & Sesudah Leetspeak.....	25
Gambar 4.3 Penerapan Frequency Analysis.....	26
Gambar 4.4 Hasil 25 Kata Paling Sering Muncul.....	27
Gambar 4.5 Hasil Bigram Paling Sering Muncul.....	27
Gambar 4.6 Hasil Trigram Paling Sering Muncul.....	27
Gambar 4.7 Hasil Akhir Seluruh Kata Kunci.....	28
Gambar 4.8 Labelling Manual Else-If.....	28
Gambar 4.9 Contoh Setelah Tokenisasi.....	29
Gambar 4.10 Data Training & Validation.....	30
Gambar 4.11. Konstruksi Arsitektur Model.....	30
Gambar 4.12. Proses Training.....	31
Gambar 4.13. Langsung Melakukan Pencocokan Setelah Frequency Analysis...32	
Gambar 4.14. Labelling Menggunakan Fuzzy Matching.....	33
Gambar 4.15. Hasil Labelling Fuzzy Matching.....	34
Gambar 4.16. Learning Curve Skenario 1 – 3.....	35
Gambar 4.17. Grafik Prediksi Model IndoBERT.....	36

Gambar 4.18. Grafik Prediksi Model Fuzzy Matching.....	38
Gambar 4.19. Hasil Fuzzy Matching Tipe 1.....	39
Gambar 4.20. Hasil Fuzzy Matching Tipe 2.....	39
Gambar 4.21. Hasil Fuzzy Matching Tipe 3.....	39
Gambar 4.22. Learning Curve Skenario 7 – 9.....	40
Gambar 4.23. Grafik Prediksi Model Fuzzy Matching.....	41
Gambar 4.24. Sampling 300 Data Tipe 1.....	42
Gambar 4.25. Sampling 300 Data Tipe 2.....	43
Gambar 4.26. Sampling 300 Data Tipe 3.....	43
Gambar 4.27. Prediksi Data Sampling Oleh Model.....	44
Gambar 4.28. Hasil Prediksi Model Dataset Tipe 1.....	44
Gambar 4.29. Hasil Prediksi Model Dataset Tipe 2.....	44
Gambar 4.30. Hasil Prediksi Model Dataset Tipe 3.....	45
Gambar 4.31. Metrik Evaluasi Sampling Tipe 1.....	46
Gambar 4.32. Metrik Evaluasi Sampling Tipe 2.....	46
Gambar 4.33. Metrik Evaluasi Sampling Tipe 3.....	46
Gambar 4.34. Kesalahan Prediksi Tipe 1.....	47
Gambar 4.35. Kesalahan Prediksi Tipe 2.....	48
Gambar 4.36. Kesalahan Prediksi Tipe 3.....	48

DAFTAR LAMPIRAN

Lampiran A Turnitin.....	56
Lampiran B Konsultasi Bimbingan.....	59
Lampiran C Potongan Kode Sistem.....	61



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA