

**ANALISIS PERFORMA ALGORITMA MACHINE LEARNING
DALAM DETEKSI PROMOSI JUDI ONLINE PADA
KOMENTAR YOUTUBE**

SKRIPSI

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



disusun oleh

HAFIZH MOHAMMAD FATHAN

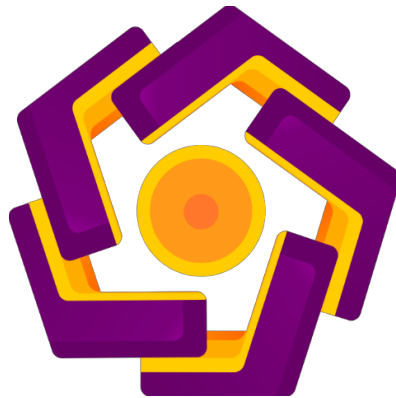
21.11.3844

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS PERFORMA ALGORITMA MACHINE LEARNING
DALAM DETEKSI PROMOSI JUDI ONLINE PADA
KOMENTAR YOUTUBE**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

HAFIZH MOHAMMAD FATHAN

21.11.3844

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS PERFORMA ALGORITMA MACHINE LEARNING DALAM
DETEKSI PROMOSI JUDI ONLINE PADA KOMENTAR YOUTUBE**

yang disusun dan diajukan oleh

Hafizh Mohammad Fathan

21.11.3844

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Juli 2025

Dosen Pembimbing,



M. Hanafi, S.Kom., M.Eng., Ph.D

NIK. 190302024

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS PERFORMA ALGORITMA MACHINE LEARNING DALAM
DETEKSI PROMOSI JUDI ONLINE PADA KOMENTAR YOUTUBE**

yang disusun dan diajukan oleh

Hafizh Mohammad Fathan

21.11.3844

Telah dipertahankan di depan Dewan Penguji
pada tanggal 29 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393



Windha Mega Pradnya Duhita, S.Kom., M.Kom.
NIK. 190302185



Hanafi, S.Kom., M.Eng., Ph.D.
NIK. 190302024



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

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, S.Kom., M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Hafizh Mohammad Fathan
NIM : 21.11.3844

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Performa Algoritma Machine Learning Dalam Deteksi Promosi Judi Online Pada Komentar Youtube

Dosen Pembimbing : M. Hanafi, S.Kom., M.Eng., Ph.D

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 29 Juli 2025

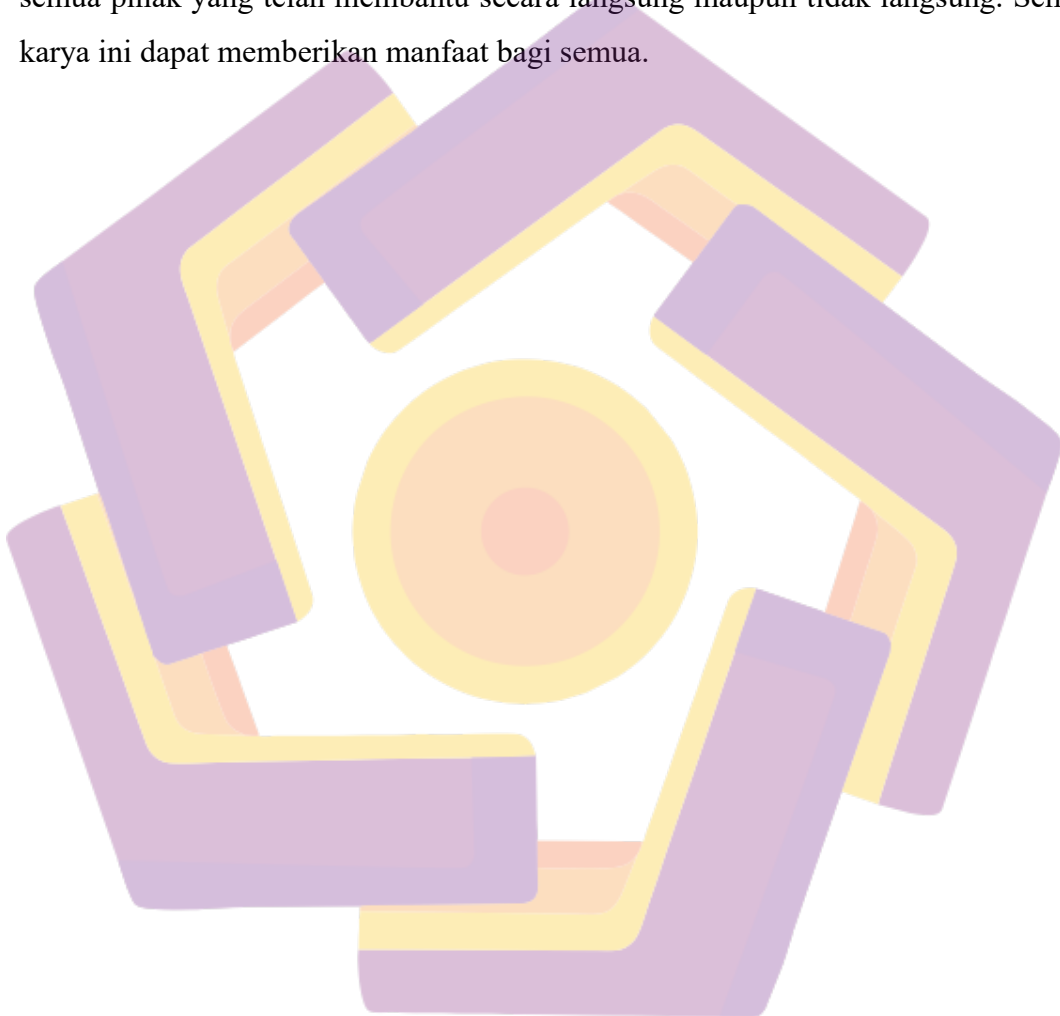
Yang Menyatakan,



Hafizh Mohammad Fathan

HALAMAN PERSEMBAHAN

Skripsi ini penulis persembahkan kepada Allah SWT atas segala rahmat dan kemudahan-Nya. Ucapan terima kasih yang mendalam penulis sampaikan kepada kedua orang tua dan keluarga besar atas doa, dukungan, dan semangat yang tiada henti. Secara khusus, penulis mempersembahkan karya ini kepada paman yang telah menjadi sumber inspirasi, motivasi, serta memberikan dukungan moral maupun materi selama masa studi. Tidak lupa, terima kasih juga penulis sampaikan kepada teman-teman dan semua pihak yang telah membantu secara langsung maupun tidak langsung. Semoga karya ini dapat memberikan manfaat bagi semua.



KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas karunia, rahmat, dan ridho-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Analisis Performa Algoritma Machine Learning Dalam Deteksi Promosi Judi Online Pada Komentar Youtube”**. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Universitas Amikom Yogyakarta.

Dalam kesempatan ini, penulis menyampaikan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Bapak M. Hanafi, S.Kom., M.Eng., Ph.D, selaku dosen pembimbing yang telah memberikan arahan, bimbingan, serta dukungan selama proses penyusunan skripsi ini
2. Kedua orang tua, keluarga besar, serta paman yang selalu memberikan doa, motivasi, dan dukungan moral maupun materi.
3. Teman-teman dan semua pihak yang telah membantu secara langsung maupun tidak langsung.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, segala kritik dan saran yang membangun sangat diharapkan. Semoga skripsi ini dapat memberikan manfaat bagi semua pihak.

Yogyakarta, 9 Juni 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL..... i

HALAMAN PERSETUJUAN..... ii

HALAMAN PENGESAHAN..... iii

HALAMAN PERNYATAAN KEASLIAN SKRIPSI iv

HALAMAN PERSEMBAHAN v

KATA PENGANTAR vi

DAFTAR TABEL xi

DAFTAR GAMBAR xii

DAFTAR LAMPIRAN..... xiii

DAFTAR LAMBANG DAN SINGKATAN xiv

DAFTAR ISTILAH xv

INTISARI..... xvi

ABSTRACT..... xvii

BAB I PENDAHULUAN..... 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 3

1.3 Batasan Masalah..... 3

1.4 Tujuan Penelitian 3

1.5 Manfaat Penelitian 4

1.6 Sistematika Penulisan 4

BAB II TINJAUAN PUSTAKA..... 7

2.1 Studi Literatur 7

2.2 Dasar Teori..... 13

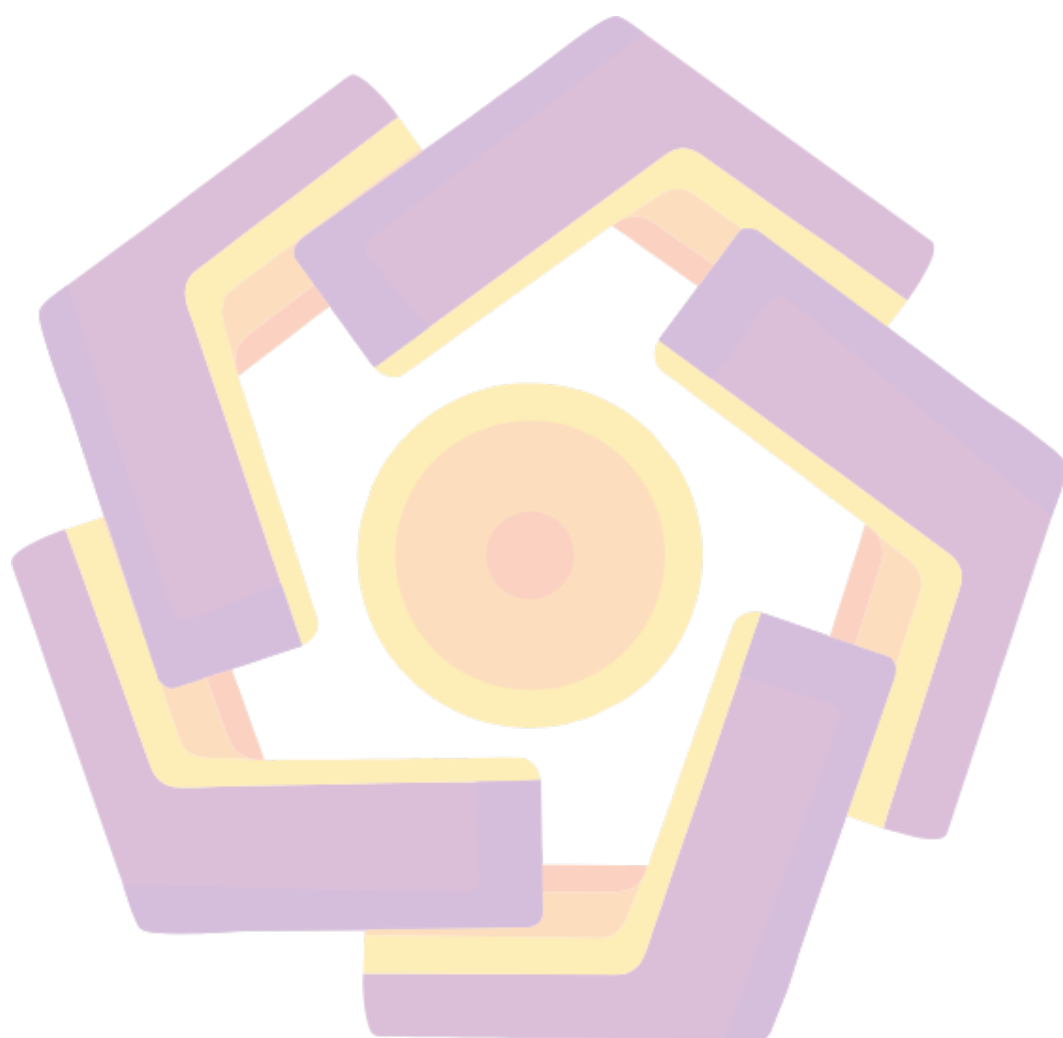
2.2.1 Judi Online 13

2.2.2 Teknologi Machine Learning 14

2.2.3 Algoritma yang Digunakan 15

2.2.3.1	Naive Bayes	15
2.2.3.2	Support Vector Machine (SVM).....	16
2.2.3.3	Random Forest	16
2.2.3.4	Logistic Regression.....	18
2.2.3.4	K-Nearest Neighbor (kNN).....	19
2.2.4	Teknik Representasi Data Teks.....	20
2.2.5	Evaluasi Model Klasifikasi Teks	21
2.2.5.1	Confusion matrix.....	22
2.2.5.2	Matrix Evaluasi (<i>Evaluation Metrics</i>).....	23
BAB III METODE PENELITIAN.....		25
3.1	Objek Penelitian.....	25
3.2	Alur Penelitian	25
3.2.1	Pengumpulan Data	26
3.2.2	Preprocessing	27
3.2.3	Labeling.....	28
3.2.4	Balancing, Tokenisasi, dan Representasi Data	28
3.2.4.1	Balancing.....	28
3.2.4.2	Tokenisasi	29
3.2.4.3	TF-IDF	29
3.2.5	Pembagian Data	30
3.2.6	Pelatihan Model	30
3.2.7	Evaluasi Performa	31
3.3	Alat dan Bahan.....	32
3.3.1	Server untuk Analisis Data.....	32
3.3.2	Laptop untuk Remote Access	33

3.3.3	Software yang Digunakan	33
3.3.4	Dataset yang Digunakan	35
3.4	Deployment	36
BAB IV HASIL DAN PEMBAHASAN		39
4.1	Hasil Penelitian	39
4.1.1	Pengumpulan data	39
4.1.2	Preprocessing	40
4.1.3	Labeling.....	42
4.1.4	Balancing, Tokenisasi, dan Representasi Data	43
4.1.5	Implementasi Model.....	44
4.1.6	Evaluasi Kinerja Algoritma.....	45
4.1.7.1	Naive Bayes	45
4.1.7.2	Support Vector Machine (SVM).....	47
4.1.7.3	Random Forest	49
4.1.7.4	k-Nearest Neighbors (kNN)	52
4.1.7.5	Logistic Regression.....	54
4.1.7.6	Evaluasi keseluruhan algoritma	56
4.1.6	Deployment.....	61
BAB V PENUTUP.....		62
5.1	Kesimpulan	62
5.2	Saran.....	64
REFERENSI		65
LAMPIRAN.....		68



DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian..... 10

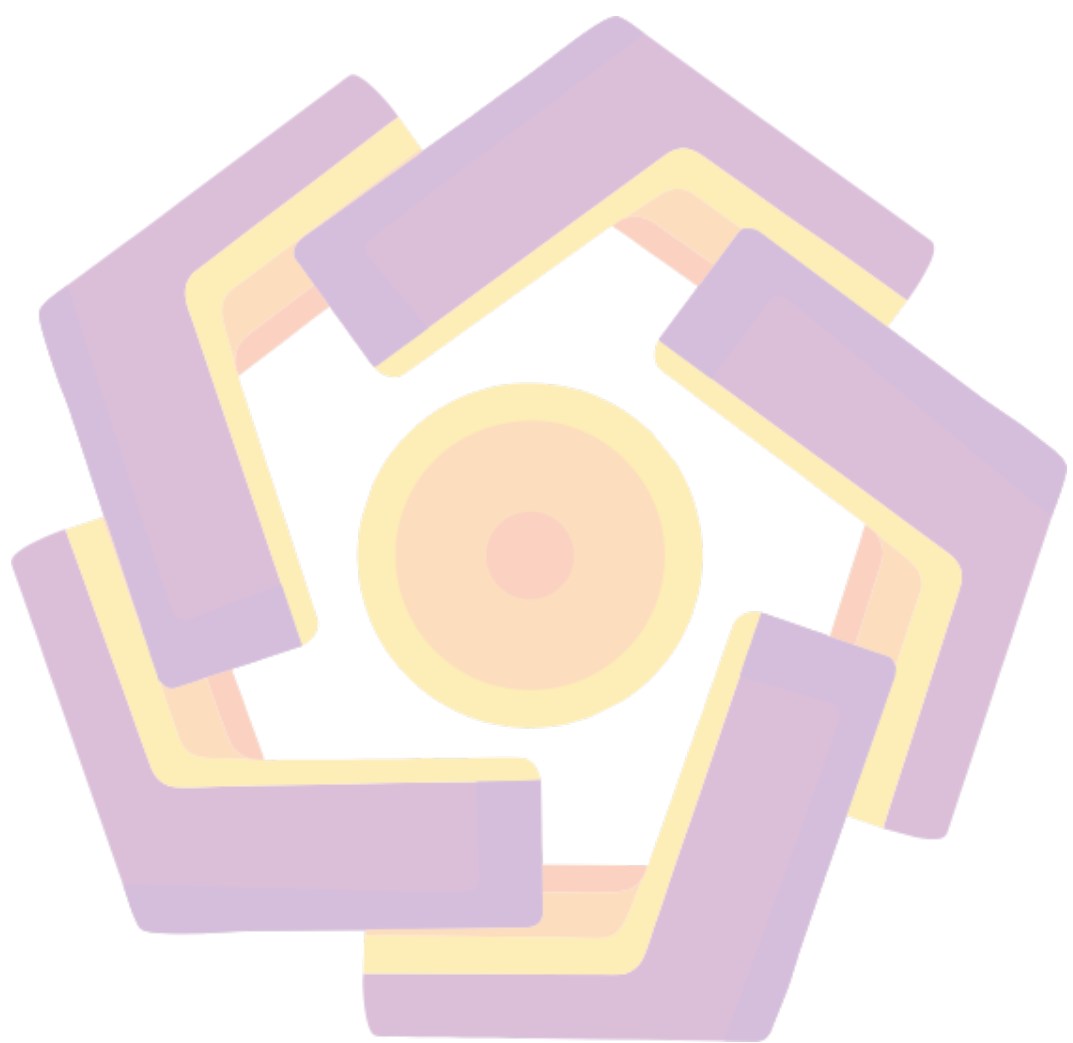
Tabel 3. 1 Spesifikasi Server..... 32

Tabel 3. 2 Spesifikasi Laptop..... 33

Tabel 3. 3 Daftar Software 33

Tabel 4. 1 pra-pemrosesan data..... 41

Tabel 4. 2 Labeling 42



DAFTAR GAMBAR

Gambar 2. 1 Confusion Matrix 22

Gambar 3. 1 Alur Penelitian..... 26

Gambar 3. 2 Skema rancangan implementasi 36

Gambar 4. 1 Isi dataset..... 39

Gambar 4. 2 informasi dataset 40

Gambar 4. 3 Analisis TF-IDF 43

Gambar 4. 4 Diagram Distribusi label 44

Gambar 4. 5 Laporan Klasifikasi Naive Bayes..... 45

Gambar 4. 6 ROC Curve Naive Bayes..... 46

Gambar 4. 7 Confusion Matrix Naive Bayes 46

Gambar 4. 8 Laporan Klasifikasi SVM..... 47

Gambar 4. 9 ROC Curve SVM 48

Gambar 4. 10 confusion Matrix SVM 49

Gambar 4. 11 Laporan Klasifikasi Random Forest..... 50

Gambar 4. 12 ROC Curve Random Forest 50

Gambar 4. 13 confusion Matrix Random Forest..... 51

Gambar 4. 14 Laporan Klasifikasi KNN 52

Gambar 4. 15 ROC Curve K-NN..... 52

Gambar 4. 16 confusion Matrix K-NN 53

Gambar 4. 17 Laporan Klasifikasi Logistic Regression 54

Gambar 4. 18 ROC Curve Logistic Regression..... 55

Gambar 4. 19 confusion Matrix Logistic Regression 56

Gambar 4. 20 Grafik Perbandingan Akurasi dan AUC..... 57

Gambar 4. 21 Grafik Perbandingan Evaluasi Class 0..... 57

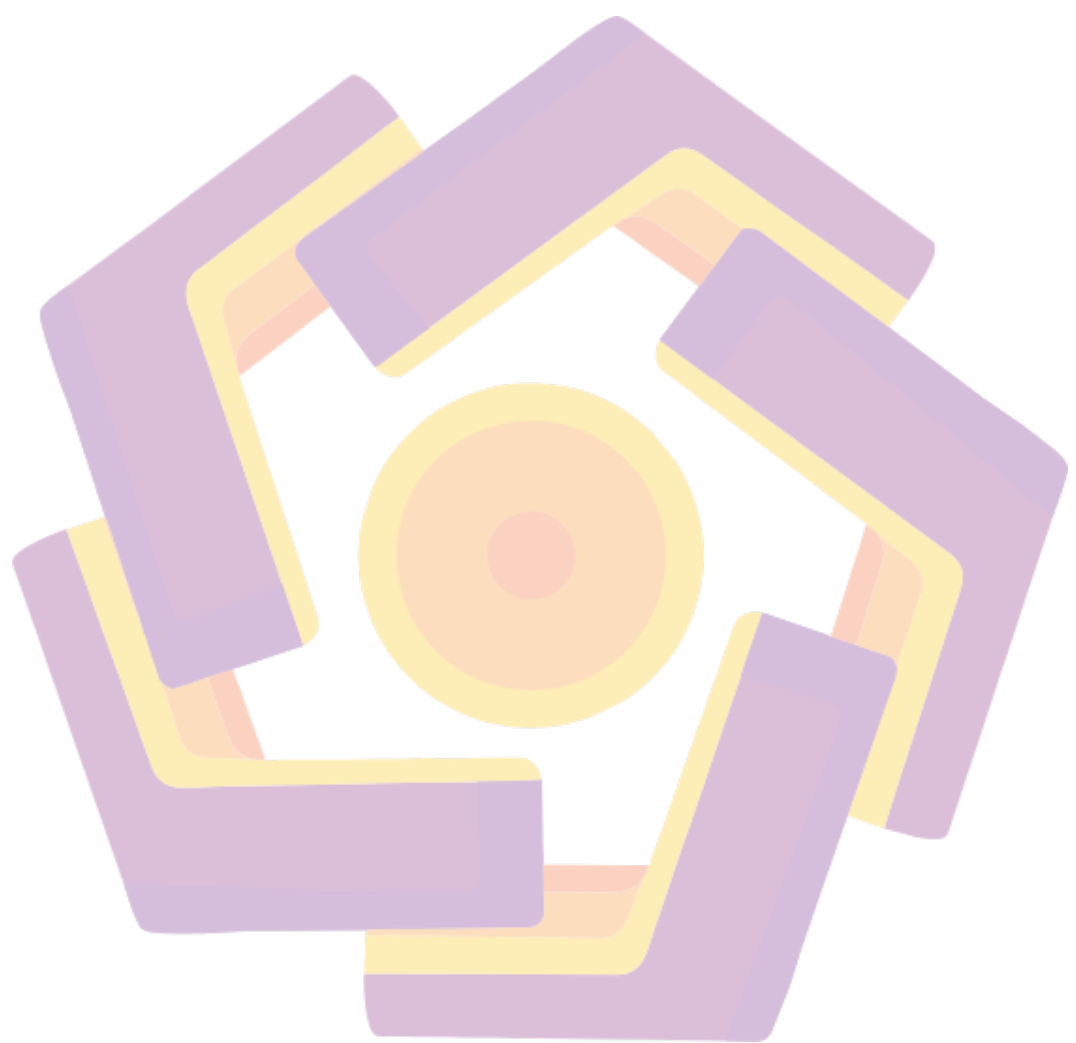
Gambar 4. 22 Grafik Perbandingan Evaluasi Class 1 58

Gambar 4. 23 Display Hasil Dashboard..... 61

DAFTAR LAMPIRAN

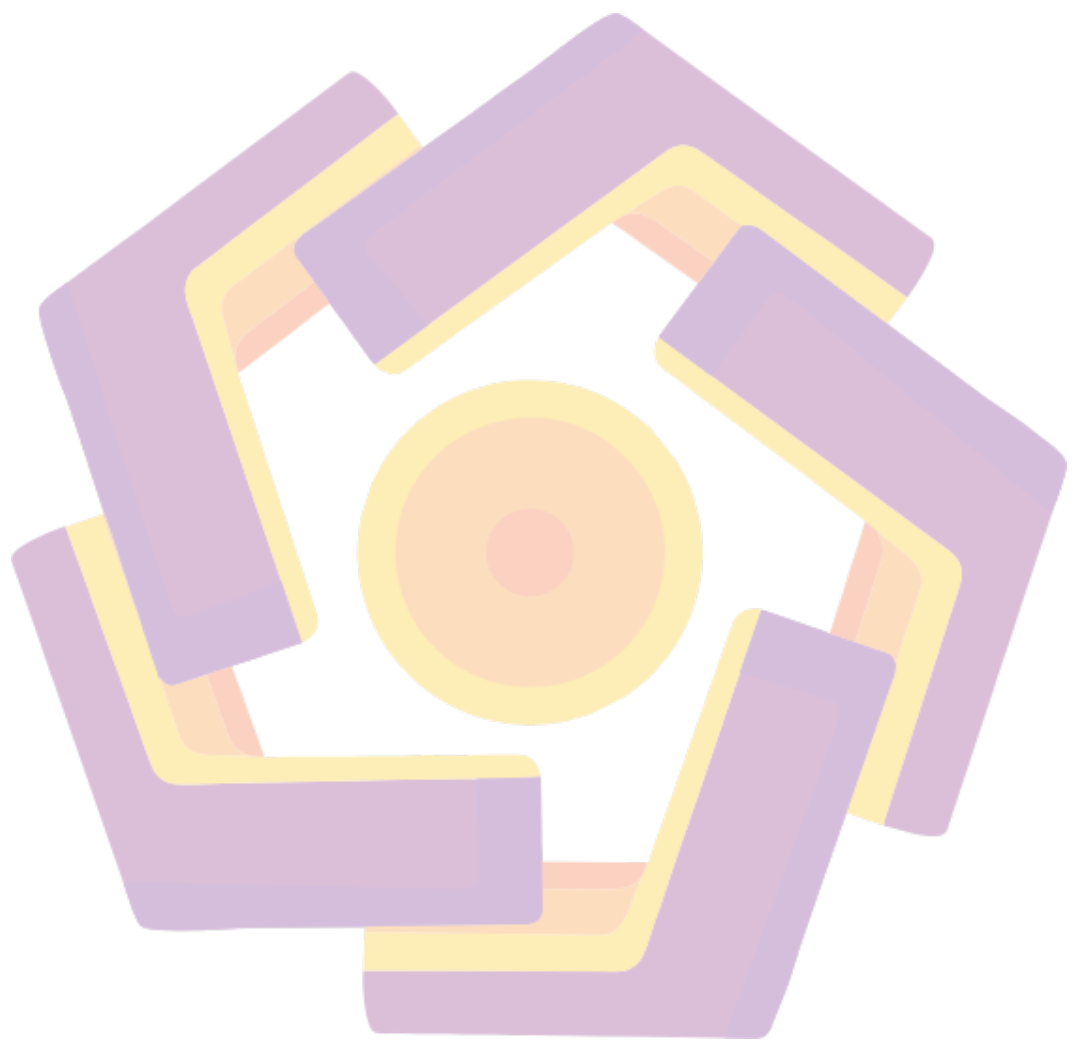
Lampiran 1. Dokumentasi Deployment 84

Lampiran 2. Dokumentasi Kode Penelitian 84



DAFTAR LAMBANG DAN SINGKATAN

AUC	<i>Area Under the Receiver Operating Characteristic Curve</i>
SVM	<i>Support Vector Machines</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
KNN	<i>K-Nearest Neighbor</i>
KUHP	Kitab Undang-Undang Hukum Pidana
ML	Machine Learning
PPATK	Pusat Pelaporan dan Analisis Transaksi Keuangan
ROC	Receiver Operating Characteristic
APJII	Asosiasi Penyelenggara Jasa Internet Indonesia
UU ITE	Undang-Undang Informasi dan Transaksi Elektronik



DAFTAR ISTILAH

Judi Online	Segala bentuk permainan judi yang dilakukan melalui internet
Scrping	Proses pengambilan data secara otomatis dari situs web menggunakan program atau script tertentu.
Machine Learning	Cabang kecerdasan buatan yang memungkinkan komputer belajar dari data untuk membuat prediksi atau keputusan.
Slang	Bahasa tidak baku atau gaul yang digunakan secara informal, sering muncul dalam komentar media sosial.
Algoritma	Sekumpulan langkah logis dan sistematis yang digunakan untuk memproses data dan menyelesaikan masalah, seperti klasifikasi.
Precision	Ukuran ketepatan model dalam memprediksi kelas positif, dihitung dari rasio prediksi positif yang benar.
Recall	Ukuran kemampuan model dalam menemukan seluruh data positif yang benar, meskipun sebagian kecil salah diklasifikasikan.
F1-Score	Nilai rata-rata harmonis antara precision dan recall, digunakan untuk menilai performa model secara seimbang.
Akurasi	Persentase total prediksi yang benar dibandingkan dengan seluruh jumlah data yang diuji.

INTISARI

Perjudian online di Indonesia terus mengalami peningkatan meskipun telah diatur secara tegas dalam Pasal 303 Kitab Undang-Undang Hukum Pidana (KUHP) dan Pasal 27 ayat (2) juncto Pasal 45 ayat (2) Undang-Undang Nomor 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE). Salah satu media yang sering dimanfaatkan untuk promosi perjudian daring adalah YouTube, dengan penyamaran melalui slang, kode, maupun tautan tersembunyi yang menysasar Generasi Z dan Milenial. Penelitian ini bertujuan menganalisis dan membandingkan performa lima algoritma *machine learning*, yaitu *Naive Bayes*, *Support Vector Machine (SVM)*, *Random Forest*, *K-Nearest Neighbor (k-NN)*, dan *Logistic Regression* dalam mendeteksi komentar promosi judi online berbahasa Indonesia pada platform YouTube. Dataset penelitian diambil dari komentar yang dikumpulkan pada periode Januari hingga Maret 2025 dan direpresentasikan menggunakan *Term Frequency Inverse Document Frequency (TF-IDF)* tanpa proses pembersihan data lanjutan. Evaluasi kinerja dilakukan menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan *Area Under Curve (AUC)*. Hasil pengujian menunjukkan seluruh algoritma memiliki akurasi di atas 0,90, dengan *Random Forest* menjadi algoritma terbaik (akurasi 0,97; *AUC* 0,99; *F1-score* mendekati sempurna), diikuti oleh *SVM* dengan *recall* tinggi (0,94). *Naive Bayes* unggul pada *recall* (0,98) namun memiliki *precision* rendah, *Logistic Regression* mencapai *AUC* sempurna (1,00), sedangkan *k-NN* mencatat kinerja terendah (akurasi 0,92; *recall* 0,89). Berdasarkan fokus untuk meminimalkan lolosnya komentar promosi judi online, *Random Forest* direkomendasikan sebagai model utama sistem moderasi otomatis, dengan *SVM* sebagai alternatif. Hasil penelitian ini diharapkan dapat mendukung pengembangan teknologi deteksi konten berbahaya yang efektif serta berkontribusi pada upaya perlindungan masyarakat dari dampak negatif perjudian online.

Kata kunci: perjudian online, YouTube, *machine learning*, *Random Forest*, deteksi konten.

ABSTRACT

Online gambling in Indonesia continues to rise despite being explicitly prohibited under Article 303 of the Indonesian Criminal Code (KUHP) and Article 27 paragraph (2) in conjunction with Article 45 paragraph (2) of Law Number 11 of 2008 on Electronic Information and Transactions (UU ITE). YouTube has emerged as one of the most frequently exploited platforms for promoting online gambling, often disguising messages through slang, codes, and hidden links targeting Generation Z and Millennials. This study aims to analyze and compare the performance of five machine learning algorithms—Naive Bayes, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (k-NN), and Logistic Regression in detecting Indonesian-language online gambling promotion comments on YouTube. The dataset was collected from YouTube comments posted between January and March 2025 and represented using Term Frequency–Inverse Document Frequency (TF-IDF) without additional data-cleaning procedures. The evaluation metrics include accuracy, precision, recall, F1-score, and Area Under Curve (AUC). The experimental results show that all algorithms achieved accuracy above 0.90, with Random Forest achieving the best performance (accuracy 0.97; AUC 0.99; near-perfect F1-score), followed by SVM with a high recall of 0.94. Naive Bayes demonstrated the highest recall (0.98) but relatively low precision, Logistic Regression achieved a perfect AUC (1.00), while k-NN recorded the lowest performance (accuracy 0.92; recall 0.89). Considering the primary objective of minimizing undetected online gambling promotional comments, Random Forest is recommended as the primary model for an automated moderation system, with SVM as a promising alternative. The findings of this study are expected to contribute to the development of effective harmful-content detection technologies and support efforts to protect the public from the negative impacts of online gambling.

Keyword: Content Detection, YouTube Comments, Online Gambling Promotion, Machine Learning, Text Classification, TF-IDF.