IMPLEMENTASI ALGORITMA RANDOM FOREST PADA MODEL DETEKSI KOMENTAR PELECEHAN SEKSUAL VERBAL BERBAHASA INDONESIA DI MEDIA SOSIAL X

SKRIPSI

LEIDEOVICO YUDHISTI
00000055683

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

IMPLEMENTASI ALGORITMA RANDOM FOREST PADA MODEL DETEKSI KOMENTAR PELECEHAN SEKSUAL VERBAL BERBAHASA INDONESIA DI MEDIA SOSIAL X

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

LEIDEOVICO YUDHISTI
00000055683

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama	:	Leideovico Yudhisti
Nomor Induk Mahasiswa	:	00000055683
Program Studi	:	Informatika

Skripsi dengan judul:

Implementasi Algoritma Random Forest pada Model Deteksi Komentar Pelecehan Seksual Verbal Berbahasa Indonesia di Media Sosial X

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, 4 Juli 2025

(Leideovico Yudhisti)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI ALGORITMA RANDOM FOREST PADA MODEL DETEKSI KOMENTAR PELECEHAN SEKSUAL VERBAL BERBAHASA INDONESIA DI MEDIA SOSIAL X

oleh

Nama	:	Leideovico Yudhisti
NIM	:	00000055683
Program Studi	:	Informatika
Fakultas	:	Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 10 Juli 2025
Pukul 15.00 s/s 17.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Dr. Ivransa Zuhdi Pane, B.Eng., M.Eng.)
NIDN: 8812520016

Penguji

(Eka Jaya Harsono, S.Kom., M.Eng.Sc.)
NIDN: 8[illegible]30333

Pembimbing

(Suwito Pomalingo, S.Kom., M.Kom.)
NIDN: 0911098201

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	:	Leidcovico Yudhisti
NIM	:	00000055683
Program Studi	:	Informatika
Jenjang	:	S1
Judul Karya Ilmiah	:	Implementasi Algoritma Random Forest pada Model Deteksi Komentar Pelecehan Seksual Verbal Berbahasa Indonesia di Media Sosial X

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

☑ 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 tiga tahun.

Tangerang, 4 Juli 2025

Yang menyatakan

Leideovico Yudhisti

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

HALAMAN PERSEMBAHAN / MOTTO

”A good name is to be more desired than great wealth, Favor is better than silver and gold.”

Proverbs 22:1 (NASB)

KATA PENGANTAR

Dengan penuh rasa syukur ke hadirat Tuhan Yang Maha Esa, penulis dapat menyelesaikan Tugas Akhir yang berjudul "Implementasi Algoritma Random Forest pada Model Deteksi Komentar Pelecehan Seksual Verbal Berbahasa Indonesia di Media Sosial X". Penelitian ini bertujuan untuk menganalisis performa algoritma Random Forest dalam mendeteksi komentar pelecehan seksual berbasis teks. Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Bapak Suwito Pomalingo, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Tiga mahasiswa Program Studi Bahasa Indonesia yang telah mendukung proses skripsi ini dalam melakukan pelabelan dataset untuk mendukung penelitian ini.
6. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini bermanfaat bagi pengembangan metode deteksi pelecehan seksual verbal di media sosial serta menjadi referensi bagi penelitian selanjutnya dalam menerapkan algoritma Random Forest untuk kasus serupa.

Tangerang, 4 Juli 2025

Leideovico Yudhisti

IMPLEMENTASI ALGORITMA RANDOM FOREST PADA MODEL DETEKSI KOMENTAR PELECEHAN SEKSUAL VERBAL BERBAHASA INDONESIA DI MEDIA SOSIAL X

Leideovico Yudhisti

ABSTRAK

Meningkatnya penyebaran komentar pelecehan seksual verbal di media sosial X menjadi salah satu bentuk Kekerasan Berbasis Gender *Online* (KBGO) yang mengancam remaja di Indonesia. Penelitian sebelumnya seringkali memiliki cakupan yang terlalu luas, seperti deteksi seksisme secara umum, dan belum secara spesifik mengukur potensi performa terbaik dari algoritma *Random Forest* pada komentar pelecehan seksual. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Random Forest* dan menemukan performa terbaiknya untuk model deteksi komentar pelecehan seksual verbal berbahasa Indonesia. Metodologi penelitian dimulai dengan pengumpulan 1.000 data komentar dari media sosial X, yang kemudian dilabeli secara manual. Untuk mengatasi masalah kelas tidak seimbang, diterapkan teknik *random undersampling* yang menghasilkan 757 data. Data tersebut kemudian melalui tahap *preprocessing* dan *feature extraction* menggunakan TF-IDF. Eksperimen dilakukan dengan membandingkan model *baseline* dengan model yang telah melalui proses *hyperparameter tuning* menggunakan *GridSearchCV*. Hasil evaluasi menunjukkan bahwa model dengan performa terbaik berhasil mencapai akurasi sebesar 87,50%. Peningkatan paling signifikan terlihat pada metrik *Recall* untuk kelas *Sexual Harassment* yang mencapai 92,86%, menunjukkan kemampuan model yang andal dalam mengidentifikasi komentar pelecehan seksual. Temuan ini menunjukkan algoritma *Random Forest* efektif untuk diimplementasikan pada model deteksi otomatis komentar pelecehan seksual verbal berbahasa Indonesia di media sosial X.

Kata kunci: Algoritma *Random Forest*, Deteksi Pelecehan Seksual, *Hyperparameter Tuning*, KBGO, Klasifikasi Teks, TF-IDF

IMPLEMENTATION OF THE RANDOM FOREST ALGORITHM IN A MODEL FOR DETECTING VERBAL SEXUAL HARASSMENT COMMENTS IN INDONESIAN ON SOCIAL MEDIA X

Leideovico Yudhisti

ABSTRACT

The increasing spread of verbal sexual harassment comments on social media X has become a form of Online Gender-Based Violence (OGBV) that threatens adolescents in Indonesia. Previous research has often been too broad in scope, such as detecting sexism in general, and has not specifically measured the optimal performance potential of effective algorithms for this particular case. This research aims to implement the Random Forest algorithm and find its best performance for a model that detects Indonesian-language verbal sexual harassment comments. The research methodology began with the collection of 1,000 comments from social media X, which were then manually labeled. To address the class imbalance problem, a random undersampling technique was applied, resulting in a dataset of 757 instances. This data then underwent text preprocessing and feature extraction using TF-IDF. The experiment was conducted by comparing a baseline model with a model optimized through a hyperparameter tuning process using GridSearchCV. Evaluation results show that the best-performing model achieved an accuracy of 87.50%. The most significant improvement was observed in the Recall metric for the Sexual Harassment class, which reached 92.86%, indicating the model's reliable ability to identify harassment cases. These findings demonstrate that a Random Forest algorithm is effective for implementing an automated detection model for verbal sexual harassment comments.

Keywords: *Gender-Based Online Violence, Hyperparameter Tuning, Random Forest Algorithm, Sexual Harassment Detection, Social Media X, Text Classification*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR KODE

DAFTAR RUMUS

DAFTAR LAMPIRAN