

**PENGEMBANGAN MODUL U-TAPIS: SISTEM DETEKSI
HUJATAN MENGGUNAKAN SUPPORT
VECTOR MACHINE**



SKRIPSI

**JEFER SETIAWAN
00000059297**

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

**PENGEMBANGAN MODUL U-TAPIS: SISTEM DETEKSI
HUJATAN MENGGUNAKAN SUPPORT
VECTOR MACHINE**



SKRIPSI

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

JEFER SETIAWAN

00000059297

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

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Jefer Setiawan

Nomor Induk Mahasiswa : 00000059297

Program Studi : Informatika

Skripsi dengan judul:

Pengembangan Modul U-TAPIS: Sistem Deteksi Hujan Menggunakan Support Vector Machine

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



(Jefer Setiawan)

UMM
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PENGEMBANGAN MODUL U-TAPIS: SISTEM DETEKSI HUJATAN MENGUNAKAN SUPPORT VECTOR MACHINE

oleh

Nama : Jefer Setiawan
NIM : 00000059297
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

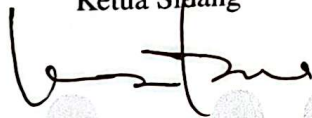
Telah diujikan pada hari Jumat, 11 Juli 2025

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

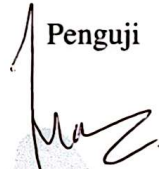
Ketua Sidang



(Dr. Ivranza Zuhdi Pane, B.Eng.,
M.Eng.)

NIDN: 8812520016

Penguji



(Marlinda Vasty Overbeek, S.Kom.,
M.Kom.)

NIDN: 0818038501

Pembimbing



(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)

NIDN: 0311106903

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 : Jefer Setiawan
NIM : 00000059297
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Pengembangan Modul U-TAPIS:
Sistem Deteksi Hujatan Menggunakan
Support Vector Machine

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

Jefer Setiawan

**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

"I can do all things through Christ who strengthens me."

Philippians 4:13 (NKJV)



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan tugas akhir yang berjudul *“Pengembangan Modul U-TAPIS: Sistem Deteksi Hujatan Menggunakan Support Vector Machine”* ini dengan baik dan tepat waktu. Dalam penyusunan tugas akhir ini, penulis mendapatkan banyak bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan terima kasih 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 Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Orang tua dan keluarga saya yang telah memberikan dukungan secara moril dan materil selama proses pengerjaan tugas akhir ini.
6. SPoT sebagai komunitas yang telah mendukung dan memotivasi saya untuk mengerjakan dan menyelesaikan tugas akhir ini.

Akhir kata, penulis berharap semoga karya ilmiah ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan di bidang teknologi informasi maupun bagi masyarakat luas, khususnya dalam mendukung upaya deteksi ujaran kebencian di ranah digital.

Tangerang, 4 Juli 2025



Jefer Setiawan

PENGEMBANGAN MODUL U-TAPIS: SISTEM DETEKSI HUJATAN MENGUNAKAN SUPPORT VECTOR MACHINE

Jefer Setiawan

ABSTRAK

Perkembangan teknologi dan penggunaan media sosial yang masif telah memunculkan berbagai tantangan baru, salah satunya adalah penyebaran ujaran kebencian atau hujatan di ruang digital. Penelitian ini bertujuan untuk mengembangkan sebuah sistem deteksi hujatan berbasis teks berbahasa Indonesia menggunakan algoritma *Support Vector Machine* (SVM). Dataset yang digunakan terdiri dari 12.000 data, yang mencakup data sintetik hujatan dari ChatGPT, data non-hujatan dari penelitian terdahulu, dan data uji yang telah divalidasi oleh ahli bahasa. Proses *preprocessing* mencakup pembersihan teks, normalisasi kata alay, dan penghapusan *stopwords*. Ekstraksi fitur dilakukan dengan menggabungkan *TF-IDF word-level*, *char-level*, dan fitur leksikon berbasis kamus hujatan. Untuk mengoptimalkan performa model, dilakukan *hyperparameter tuning* menggunakan pustaka Optuna dengan pendekatan *Bayesian Optimization*. Hasil evaluasi menunjukkan bahwa model terbaik yang dihasilkan memiliki akurasi sebesar 86%, dengan nilai *precision* rata-rata 0.86, *recall* rata-rata 0.86, dan *F1-score* rata-rata 0.86 pada setiap kelas. Sistem ini diimplementasikan dalam bentuk aplikasi berbasis web yang memungkinkan pengguna menguji kalimat untuk mendeteksi potensi hujatan secara langsung.

Kata kunci: Deteksi Hujatan, Kamus Hujatan, Optuna, Support Vector Machine, TF-IDF.



**DEVELOPMENT OF THE U-TAPIS MODULE: HATE SPEECH
DETECTION SYSTEM USING SUPPORT VECTOR MACHINE**

Jefer Setiawan

ABSTRACT

The rapid advancement of technology and the widespread use of social media have given rise to new challenges, one of which is the spread of hate speech or offensive language in digital spaces. This study aims to develop a text-based hate speech detection system for the Indonesian language using the Support Vector Machine (SVM) algorithm. The dataset consists of 12,000 entries, including synthetic hate speech generated by ChatGPT, non-hate speech data from previous research, and test data validated by linguists. The preprocessing stage involves text cleaning, normalization of slang words, and stopword removal. Feature extraction combines TF-IDF at the word and character levels, as well as lexicon-based features derived from a hate speech dictionary. To optimize model performance, hyperparameter tuning is conducted using the Optuna library with a Bayesian Optimization approach. Evaluation results show that the best-performing model achieves an accuracy of 86%, with an average precision, recall, and F1-score of 0.86 across all classes. The system is implemented as a web-based application that allows users to test sentences for potential hate speech in real time.

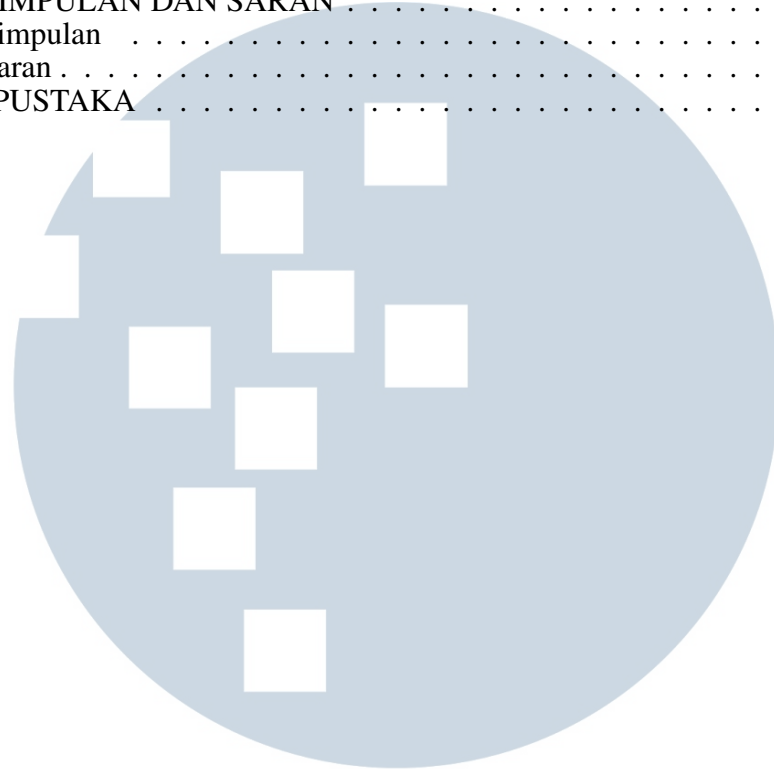
Keywords: Hate Speech Detection, Hate Word Lexicon, Optuna, Support Vector Machine, TF-IDF.



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR RUMUS	xiii
DAFTAR LAMPIRAN	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB 2 LANDASAN TEORI	6
2.1 Hate Speech dan Harassment	6
2.2 Text Classification	6
2.3 Dataset dan Validitas Data	7
2.4 Data Sampling	8
2.5 Support Vector Machine	9
2.6 Feature Extraction dengan TF-IDF Vectorizer	12
2.7 Evaluasi Kinerja Model	15
2.8 Hyperparameter Tuning	17
BAB 3 METODOLOGI PENELITIAN	18
3.1 Studi Literatur dan Tinjauan Teori	19
3.2 Pengumpulan dan Persiapan Data	20
3.3 Perancangan Model Deteksi Hujatan	22
3.3.1 Preprocessing	23
3.3.2 Ekstraksi Fitur	24
3.3.3 Pelatihan dan Evaluasi Model	26
3.3.4 Optimasi Model dengan Hyperparameter Tuning	27
3.3.5 Evaluasi Lanjutan Model Setelah Tuning	28
3.4 Implementasi Model Pembelajaran Mesin Ke Aplikasi Website	29
BAB 4 HASIL DAN DISKUSI	30
4.1 Hasil dan Evaluasi Model Awal	30
4.2 Tuning Hyperparameter dan Evaluasi Lanjutan	32
4.3 Pengaruh Hyperparameter Tuning	34
4.3.1 Perbandingan Kinerja Model Sebelum dan Sesudah Tuning	35
4.3.2 Analisis Dampak Tuning terhadap Komponen Model	35
4.4 Implementasi Website	37
4.4.1 Arsitektur Sistem	38
4.4.2 Alur Proses Sistem	38

4.4.3	Contoh Implementasi	39
4.5	Hasil dan Analisis Akhir	40
BAB 5	SIMPULAN DAN SARAN	42
5.1	Simpulan	42
5.2	Saran	43
DAFTAR PUSTAKA		45



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1	Jenis-jenis Kernel yang Umum Digunakan dalam SVM . .	10
Tabel 2.2	Perbandingan TF-IDF dengan metode main dalam klasifikasi teks	14
Tabel 2.3	Confusion Matrix	16
Tabel 3.1	Komposisi Data Berdasarkan Kelas	21
Tabel 4.1	Classification report model awal	31
Tabel 4.2	Classification report model setelah tuning	34
Tabel 4.3	Perbandingan performa model sebelum dan sesudah tuning	35
Tabel 4.4	Teknologi yang digunakan dalam implementasi website . .	38



DAFTAR GAMBAR

Gambar 2.1	Ilustrasi Kernel Trick: pemetaan data dari ruang dua dimensi ke tiga dimensi untuk mencapai separabilitas linear	10
Gambar 2.2	Ilustrasi cara kerja support vector machine	12
Gambar 3.1	Flowchart metode penelitian	18
Gambar 3.2	Flowchart perancangan model	23
Gambar 3.3	Diagram ekstraksi fitur	25
Gambar 4.1	Arsitektur Pipeline Klasifikasi Teks Deteksi Hujan	30
Gambar 4.2	Confusion matrix pada model awal	31
Gambar 4.3	Confusion matrix pada model setelah dituning	34
Gambar 4.4	Perbandingan performa model sebelum dan sesudah tuning berdasarkan metrik evaluasi utama	37
Gambar 4.5	Tampilan pada frontend	40



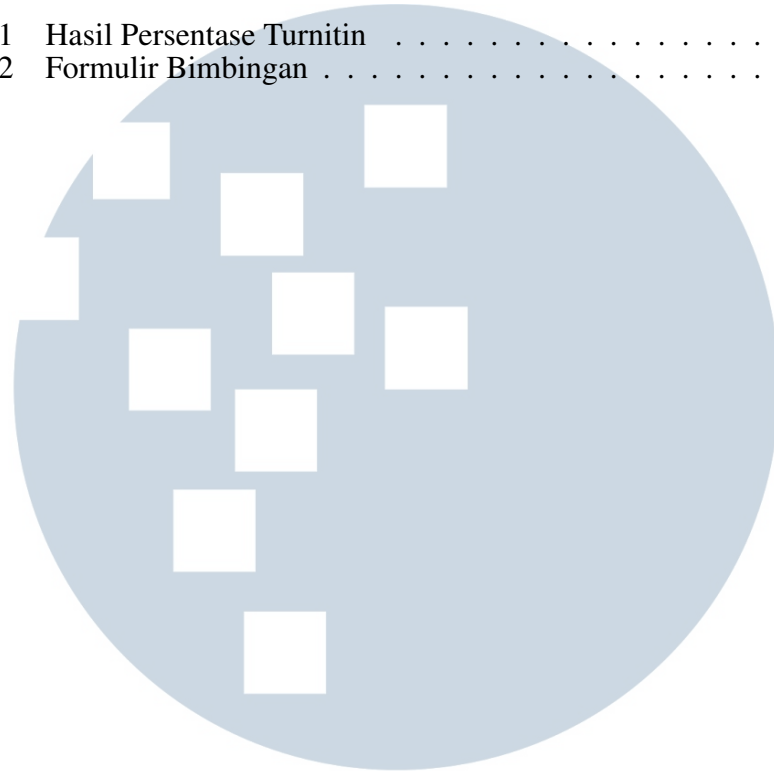
DAFTAR RUMUS

Rumus 2.1	Krejcic & Morgan	8
Rumus 2.2	Support vector machine	11
Rumus 2.3	Soft margin support vector machine	11
Rumus 2.4	<i>Term Frequency</i>	14
Rumus 2.5	<i>Inverse Document Frequency</i>	14
Rumus 2.6	<i>Term Frequency-Inverse Document Frequency</i>	15
Rumus 2.7	Rumus Akurasi	15
Rumus 2.8	Rumus Precision	15
Rumus 2.9	Rumus Recall	16
Rumus 2.10	Rumus F1-score	16



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	48
Lampiran 2	Formulir Bimbingan	56



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA