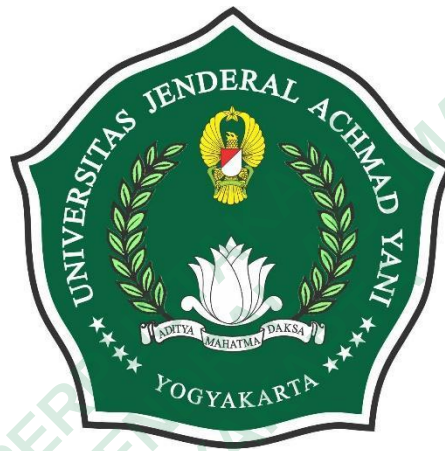


**PENERAPAN TF-IDF DALAM DETEKSI CYBERBULLYING DENGAN
ALGORITMA SUPPORT VECTOR MACHINE (SVM)**

TUGAS AKHIR

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi S-1 Teknologi Informasi



Disusun oleh:

BELA DWI UTAMI

212104003

**PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS TEKNIK & TEKNOLOGI INFORMASI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
2025**

HALAMAN PENGESAHAN

TUGAS AKHIR

**PENERAPAN TF-IDF DALAM DETEKSI *CYBERBULLYING* DENGAN
ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)**

Diajukan oleh:

BELA DWI UTAMI

212104003

Telah dipertahankan di depan dewan penguji dan dinyatakan sah
sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta

Tanggal: 16 Juli 2025

Mengesahkan:

Pembimbing

Penguji

Arief Ikhwan Wicaksono., S.Kom., M.Cs.

NIDN: 0512128401

Rama Sahtyawan, S.T., M.Cs.

NIDN: 0518058001

Ketua Program Studi S-1 Teknologi Informasi
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta



Rama Sahtyawan, S.T., M.Cs.

NPP: 2019.13.0150

PERNYATAAN

Saya yang bertanda tangan di bawah ini, adalah mahasiswa Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta,

Nama : Bela Dwi Utami
NPM : 212104003
Program Studi : S-1 Teknologi Informasi
Judul Tugas Akhir : Penerapan TF-IDF dalam Deteksi *Cyberbullying* dengan Algoritma *Support Vector Machine* (SVM)

Menyatakan bahwa hasil penelitian dengan judul tersebut di atas adalah asli karya saya sendiri dan bukan hasil plagiarisme. Semua referensi dan sumber terkait yang dikutip dalam karya ilmiah ini telah ditulis sesuai kaidah penulisan ilmiah yang berlaku. Dengan ini, saya menyatakan untuk menyerahkan hak cipta penelitian kepada Universitas Jenderal Achmad Yani Yogyakarta guna kepentingan ilmiah.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya tanpa ada paksaan dari pihak mana pun. Apabila terdapat kekeliruan atau ditemukan adanya pelanggaran akademik di kemudian hari, maka saya bersedia menerima konsekuensi yang berlaku sesuai ketentuan akademik.

Yogyakarta, 16 Juli 2025



Bela Dwi Utami

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Penerapan TF-IDF dalam Deteksi *Cyberbullying* dengan Algoritma *Support Vector Machine* (SVM)”. Penyusunan laporan ini merupakan salah satu persyaratan untuk menyelesaikan studi di Program Studi S-1 Teknologi Informasi Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Laporan ini dapat diselesaikan atas bimbingan, arahan, dan bantuan dari berbagai pihak. Pada kesempatan ini penulis dengan rendah hati mengucapkan terima kasih dengan setulus-tulusnya kepada:

1. Bapak Rama Sahtyawan, S.T., M.Cs. selaku Ketua Program Studi S-1 Teknologi Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
2. Bapak Arief Ikhwan Wicaksono., S.Kom., M.Cs. selaku Dosen Pembimbing Tugas Akhir yang telah mendukung dan membantu selama berjalannya penelitian hingga penyusunan laporan;
3. Bapak Adkhan Sholeh, S.Si., M.Cs. selaku Dosen Pembimbing Akademik;
4. Para dosen Teknologi Informasi yang telah memberikan banyak bekal ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
5. Kedua Orang tua tercinta, sesuai harapan bapak ingin melihat anaknya sarjana walaupun akhirnya berjuang tanpa beliau. Semoga Bapak bangga melihat pencapaian ini dari surga. Dan teruntuk Ibunda perempuan hebat yang menjadi tulang punggung keluarga, terima kasih atas segala doa, dukungan, dan cinta yang tak pernah berhenti sampai hari ini. Sehat selalu bu, ibu harus ada di setiap pencapaian dan perjalanan selanjutnya;
6. Kakak penulis yang selalu memberikan dukungan moril dan materil, yang memotivasi dan mendoakan penulis;
7. Kepada sahabat penulis, terimakasih telah mendengarkan keluh kesah penulis, dan selalu memberikan semangat dan dukungan kepada penulis

sehingga dapat menyelesaikan Tugas Akhir ini dengan tepat waktu, dan semua pihak yang tidak dapat penulis cantumkan satu persatu yang selalu memberikan dukungan positif selama ini;

8. Rekan-rekan mahasiswa Prodi S-1 Teknologi Informasi di Universitas Jenderal Achmad Yani Yogyakarta yang sudah memberi dukungan dan kerja sama selama pembuatan tugas akhir;
9. Dan terakhir, untuk diri saya sendiri, terimakasih sudah menepikan ego dan memilih untuk kembali bangkit dan menyelesaikan apa yang sudah dimulai. Terimakasih telah mengendalikan diri sendiri dari berbagai tekanan di luar keadaan dan memilih untuk bertahan sejauh ini.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Maka dari itu dengan segala kerendahan hati penulis sangat menghargai adanya kritik dan saran yang membangun dari semua pihak yang bersedia meluangkan waktu untuk membaca laporan tugas akhir ini.

Yogyakarta, 16 Juli 2025

Bela Dwi Utami

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
DAFTAR SINGKATAN.....	xi
INTISARI	xii
ABSTRACT	xiii
Bab 1 PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	3
1.3 Pertanyaan Penelitian	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Hasil Penelitian	4
Bab 2 Tinjauan Pustaka dan Landasan Teori.....	5
2.1 Tinjauan Pustaka	5
2.2 Landasan Teori.....	11
2.2.1 Cyberbullying	11
2.2.2 Text Mining	12
2.2.3 Machine Learning	12
2.2.4 Natural Language Processing.....	12
2.2.5 Term Frequency-Inverse Document Frequency.....	13
2.2.6 Support Vector Machine.....	14
2.2.7 Confusion matrix	15
2.2.8 Analisis False Positive dan Evaluasi Efektivitas Model.....	17

Bab 3 Metode Penelitian	19
3.1 Bahan dan Alat Penelitian.....	20
3.2 Jalan Penelitian.....	21
3.2.1 Pengumpulan Data	21
3.2.2 <i>Pre-processing</i>	21
3.2.3 Ekstraksi Fitur dengan TF-IDF	23
3.2.4 Metode <i>Baseline Centroid-Based Classifier</i>	24
3.2.5 <i>Support Vector Machine</i>	25
3.2.6 Evaluasi Model.....	26
Bab 4 Hasil Penelitian	27
4.1 Ringkasan Hasil Penelitian	27
4.2 Pembahasan Hasil Penelitian	27
4.2.1 Persiapan Dataset	27
4.2.2 <i>Pre-processing</i>	29
4.2.3 <i>Splitting</i> Data.....	38
4.2.4 Ekstraksi Fitur menggunakan TF-IDF	38
4.2.5 Klasifikasi Teks <i>Cyberbullying</i>	41
4.2.6 Evaluasi Model.....	43
Bab 5 Kesimpulan dan Saran.....	53
5.1 Kesimpulan	53
5.2 Saran.....	54
DAFTAR PUSTAKA	55
LAMPIRAN.....	59

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terdahulu.....	8
Tabel 2.2 <i>Confusion Matrix</i>	15
Tabel 4.1 Atribut Dataset	28
Tabel 4.2 Contoh Komentar pada Dataset.....	28
Tabel 4.3 Hasil Teks <i>Cleaning</i>	30
Tabel 4.4 Hasil <i>Case Folding</i>	32
Tabel 4.5 Hasil Normalisasi Teks	33
Tabel 4.6 Hasil <i>Remove Stopword</i>	35
Tabel 4.7 Hasil Lemmatization	36
Tabel 4.8 Distribusi Data Latih dan Data Uji.....	38
Tabel 4.9 <i>Classification report</i> Model <i>Cosine similarity</i>	43
Tabel 4.10 <i>Classification Report</i> Model SVM.....	45
Tabel 4.11 Perbandingan Kinerja Model.....	47

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	19
Gambar 3.2 Distribusi Data	21
Gambar 3.3 Tahapan <i>Pre-processing</i>	22
Gambar 3.4 Konsep <i>Cosine similarity</i>	25
Gambar 3.5 <i>Visualisasi SVM</i>	26
Gambar 4.1 Hasil <i>Pre-processing</i>	37
Gambar 4.2 Hasil TF-IDF <i>Vectorizer</i>	40
Gambar 4.3 <i>Heatmap Matrix</i> TF-IDF	40
Gambar 4.4 <i>Visualisasi Confusion Matrix Baseline</i>	44
Gambar 4.5 <i>Visualisasi Confusion Matrix SVM</i>	46
Gambar 4.6 Grafik Perbandingan Kompleksitas Waktu	49
Gambar 4.7 Grafik Perbandingan Akurasi	49
Gambar 4.8 Kurva ROC <i>Baseline</i>	51
Gambar 4.9 Kurva ROC SVM	51

DAFTAR LAMPIRAN

Lampiran 1 Script Pemrograman	59
Lampiran 2 Dataset dan Hasil Pra-pemrosesan Teks	64
Lampiran 3 Jadwal Penelitian	69
Lampiran 4 Lembar Bimbingan Dosen	70
Lampiran 5 Hasil Cek Plagiarisme	71

PERPUSTAKAAN
UNIVERSITAS JENDRAL ACHMAD YANI
YOGYAKARTA

DAFTAR SINGKATAN

AUC	<i>Area Under Curve</i>
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
CBOW	<i>Continuous Bag of Words</i>
CSV	<i>Comma-Separated Values</i>
C-SVC	<i>C-Support Vector-Inverse Document Frequency</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FPR	<i>False Positive Rate</i>
IDF	<i>Inverse Document Frequency</i>
IWF	<i>Inverse Word Frequency</i>
KNN	<i>K-Nearest Neighbors</i>
NLP	<i>Natural Language Processing</i>
NLTK	<i>Natural Language Toolkit</i>
Nu-SVC	<i>Nu-Support Vector Machine</i>
OOV	<i>Out of Vocabulary</i>
POS	<i>Part of Speech</i>
RBF	<i>Radial Basis Function</i>
ROC	<i>Receiver Operating Characteristic</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
SVM	<i>Support Vector Machine</i>
TF	<i>Term Frequency</i>
TF-IDF	<i>Term Frequency-Inverse Document Frequency</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
URL	<i>Uniform Resource Locator</i>