

**KOMPARASI ALGORITMA KNN DAN SVM UNTUK
DETEKSI PELECEHAN SEKSUAL VERBAL PADA
KOMENTAR MEDIA SOSIAL TIKTOK**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

PENI FEBRIAN KRISTAMI

22.12.2292

Kepada

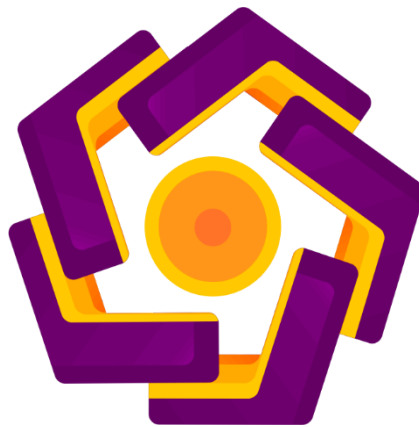
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**KOMPARASI ALGORITMA KNN DAN SVM UNTUK
DETEKSI PELECEHAN SEKSUAL VERBAL PADA
KOMENTAR MEDIA SOSIAL TIKTOK**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

PENI FEBRIAN KRISTAMI

22.12.2292

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI ALGORITMA KNN DAN SVM UNTUK DETEKSI
PELECEHAN SEKSUAL VERBAL PADA KOMENTAR MEDIA SOSIAL
TIKTOK**

yang disusun dan diajukan oleh

Peni Febrian Kristami

22.12.2292

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Februari 2026

Dosen Pembimbing,



Anggit Dwi Hartanto, S.Kom., M.Kom
NIK. 190302163

HALAMAN PENGESAHAN

SKRIPSI

**KOMPARASI ALGORITMA KNN DAN SVM UNTUK DETEKSI
PELECEHAN SEKSUAL VERBAL PADA KOMENTAR MEDIA SOSIAL
TIKTOK**

yang disusun dan diajukan oleh

Peni Febrian Kristami

22.12.2292

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Irma Rofni Wulandari, S.Pd., M.Eng.
NIK. 190302329

Dina Maulina, S.Kom., M.Kom.
NIK. 190302250

Anggit Dwi Hartanto, S.Kom., M.Kom.
NIK. 190302163

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Februari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Peni Febrian Kristami
NIM : 22.12.2292

Menyatakan bahwa Skripsi dengan judul berikut:

Komparasi Algoritma KNN dan SVM untuk Deteksi Pelecehan Seksual Verbal pada Komentar Media Sosial TikTok

Dosen Pembimbing : Anggit Dwi Hartanto, S.Kom., M.Kom

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, 24 Februari 2026

Yang Menyatakan,



Peni Febrian Kristami

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur dan kerendahan hati kepada Allah SWT, Tuhan Yang Maha Esa atas limpahan rahmat, hidayah, dan kekuatan yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Karya ini juga penulis persembahkan kepada kedua orang tua tercinta dan tersayang bapak dan ibu, yang senantiasa memberikan cinta, dukungan, semangat, dan do'a tanpa henti. Ucapan terima kasih penulis sampaikan kepada dosen pembimbing bapak Anggit Dwi Hartanto, M.Kom dan seluruh dosen yang telah memberikan arahan dan bimbingan, serta ilmu yang bermanfaat selama masa perkuliahan hingga proses penyusunan skripsi ini. Tidak lupa, penulis juga mempersembahkan karya ini kepada teman teman seperjuangan SI01 dan sahabat – sahabat tercinta Amanda, Dhinda, Edelweiss, Farikha, Iqro, dan teman - teman yang tidak dapat saya sebutkan satu per satu yang selalu memberikan semangat, kebersamaan, serta dorongan moral selama proses perjuangan ini berlangsung. Ucapan terima kasih juga penulis sampaikan kepada diri sendiri yang telah bertahan selama ini dalam menghadapi berbagai tantangan hingga skripsi ini dapat terselesaikan.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat, hidayah, serta karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Komparasi Algoritma KNN dan SVM untuk Deteksi Pelecehan Seksual Verbal pada Komentar Media Sosial TikTok” dengan lancar. Tugas akhir ini disusun untuk memenuhi salah satu syarat kelulusan Program Sarjana pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selain itu penulis dengan segala kerendahan hati ingin menyampaikan rasa terima kasih kepada semua pihak yang telah berjasa memberikan dukungan dan bantuan untuk menyelesaikan laporan tugas akhir ini. Untuk itu penulis mengucapkan terima kasih kepada :

1. Prof. Dr. M Suyanto, MM. selaku Rektor Universitas Amikom Yogyakarta
2. Hanif al fatta, S.Kom., M.Kom.Ph.D selaku Dekan Program Fakultas Ilmu Komputer
3. Anggit Dwi Hartanto, M.Kom selaku Ketua Program Studi Sistem Informasi dan dosen pembimbing yang memberikan arahan, saran, dan motivasi terhadap penulis
4. Kedua orang tua, kakak, dan adik yang senantiasa memberikan cinta, dukungan, semangat, dan do'a tanpa henti dalam setiap langkah.
5. Sahabat dan teman-teman Sistem Informasi 01 angkatan 22 atas semangat, kebersamaan, serta dorongan moral selama proses perjuangan ini berlangsung selama dibangku perkuliahan ini.

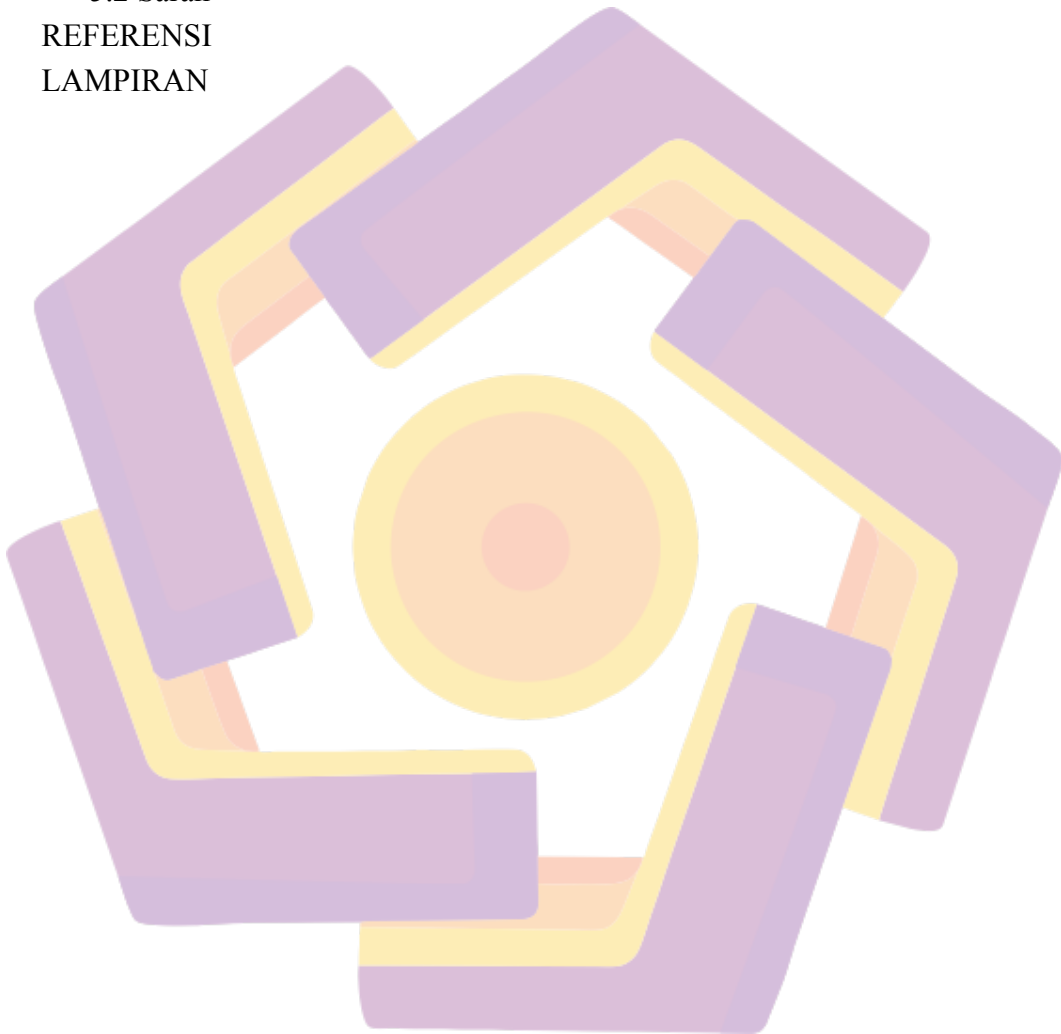
Yogyakarta, 27 Februari 2026

Penulis

DAFTAR ISI

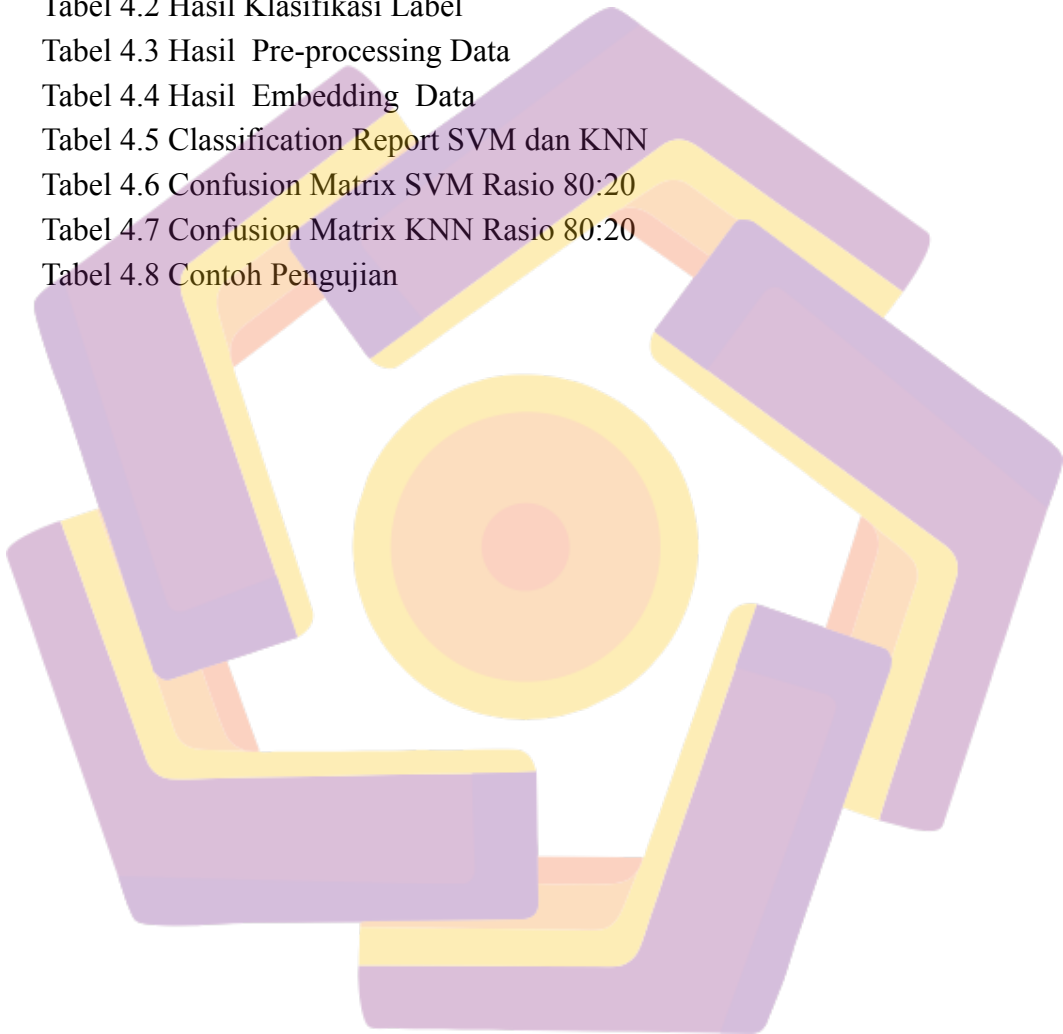
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I	
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II	
TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	18
BAB III	
METODE PENELITIAN	28
3.1 Objek Penelitian	28
3.2 Alur Penelitian	28
3.3 Alat dan Bahan	33
BAB IV	
HASIL DAN PEMBAHASAN	35
4.1 Hasil Pengumpulan Data	35
4.2 Pre Processing Data	37
4.3 Feature Extraction	40

4.4 Model Training	44
4.5 Evaluasi Model	46
4.6 Deploy Model	53
BAB V	
PENUTUP	56
5.1 Kesimpulan	56
5.2 Saran	56
REFERENSI	58
LAMPIRAN	61



DAFTAR TABEL

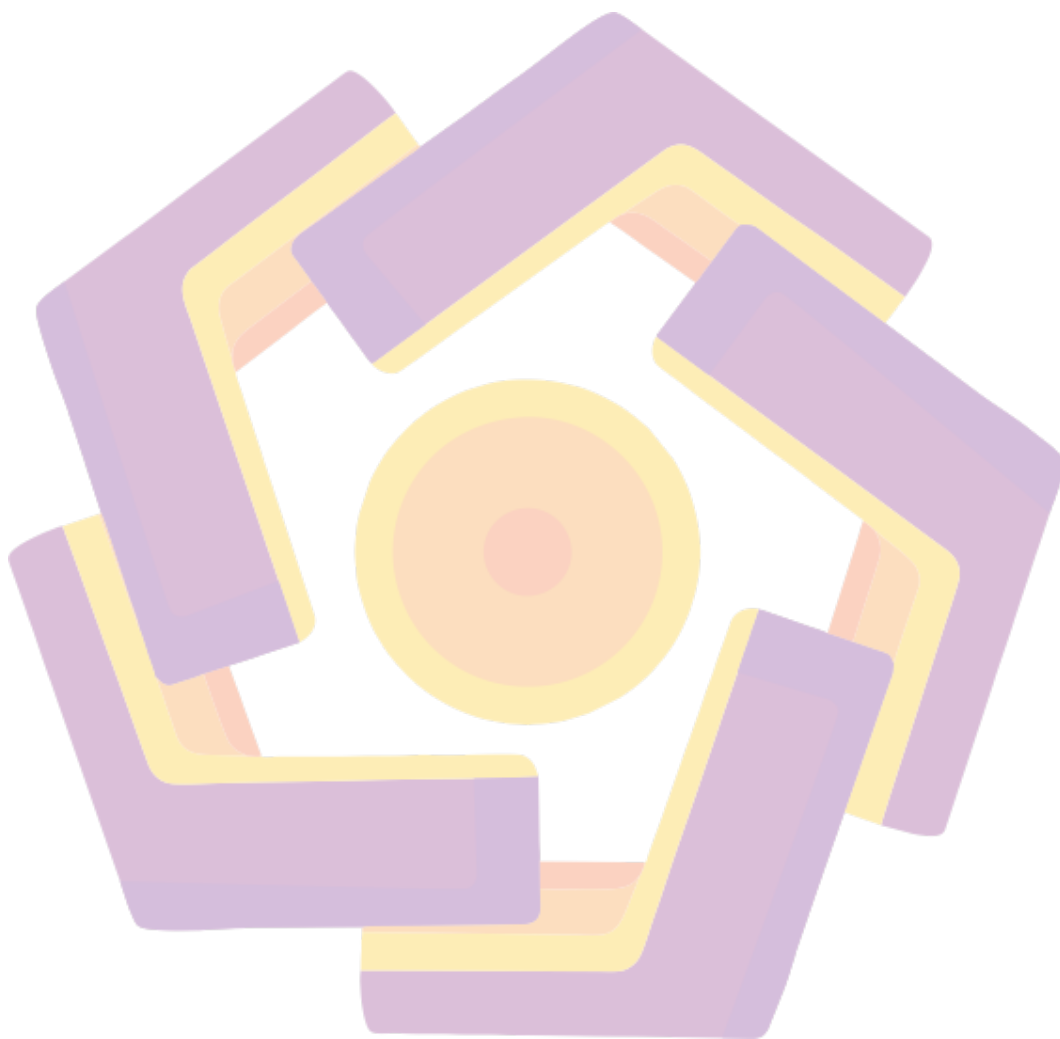
Tabel 2.1 Keaslian Penelitian	14
Tabel 4.1 Hasil Pengumpulan data	35
Tabel 4.2 Hasil Klasifikasi Label	36
Tabel 4.3 Hasil Pre-processing Data	40
Tabel 4.4 Hasil Embedding Data	42
Tabel 4.5 Classification Report SVM dan KNN	47
Tabel 4.6 Confusion Matrix SVM Rasio 80:20	49
Tabel 4.7 Confusion Matrix KNN Rasio 80:20	50
Tabel 4.8 Contoh Pengujian	52



DAFTAR GAMBAR

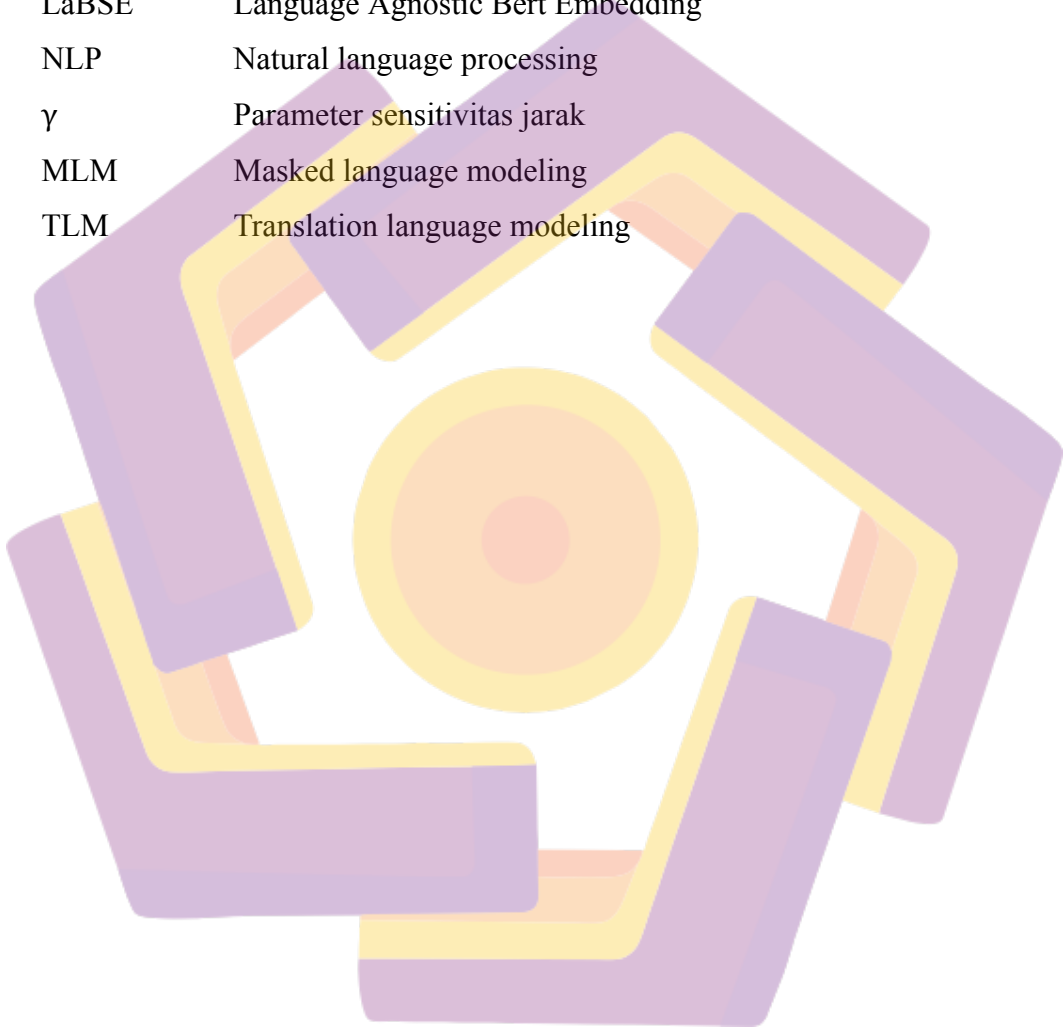
Gambar 3.1 Alur Penelitian	29
Gambar 4.1 Word Cloud Komentar Pelecehan Seksual	36
Gambar 4.2 Word Cloud Komentar Non-Pelecehan Seksual	37
Gambar 4.3 Cek Label Kosong	38
Gambar 4.4 Hasil Cleaning Dataset	38
Gambar 4.5 Label Setelah Pre-Processing	39
Gambar 4.6 Hasil Pre-processing Data	40
Gambar 4.7 Visualisasi Scatter Plot Data Embedding LaBSE	42
Gambar 4.9 Grafik Classification Report KNN	49
Gambar 4.10 Contoh Pengujian	52
Gambar 4.11 Tampilan Hasil Deploy Model	54
Gambar 4.12 Tampilan Hasil Deploy Model Positif	54
Gambar 4.13 Tampilan Hasil Deploy Model Negatif	55

DAFTAR LAMPIRAN



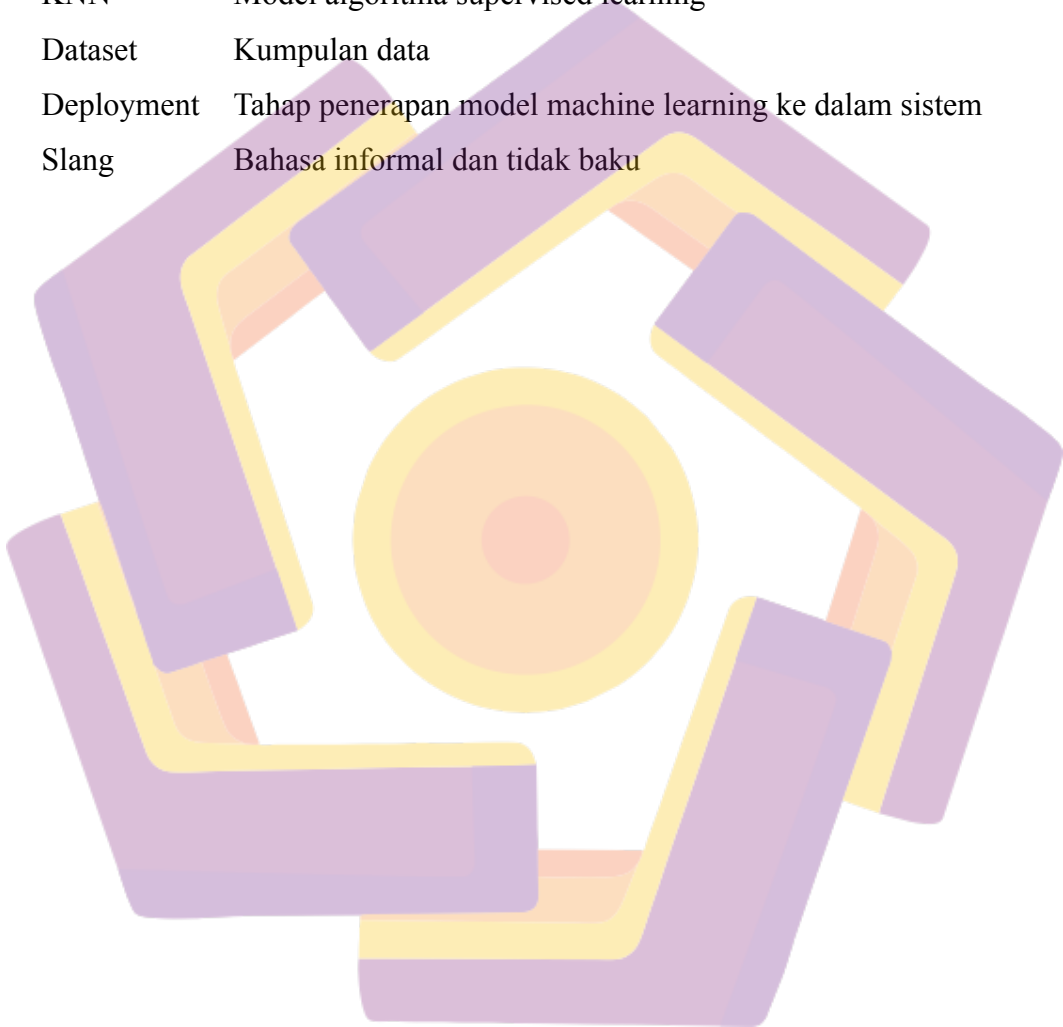
DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machines
KNN	K-Nearest Neighbor
LaBSE	Language Agnostic Bert Embedding
NLP	Natural language processing
γ	Parameter sensitivitas jarak
MLM	Masked language modeling
TLM	Translation language modeling



DAFTAR ISTILAH

NLP	Bidang ilmu yang mempelajari interaksi antara komputer dan bahasa manusia
SVM	Model algoritma supervised learning
KNN	Model algoritma supervised learning
Dataset	Kumpulan data
Deployment	Tahap penerapan model machine learning ke dalam sistem
Slang	Bahasa informal dan tidak baku



INTISARI

Pelecehan seksual secara verbal merupakan salah satu bentuk kekerasan yang banyak ditemukan di media sosial, termasuk pada kolom komentar pengguna di platform Tik Tok. Intensitas komentar yang tinggi dengan karakteristik komentar yang singkat, tidak baku, serta sering disampaikan secara implisit menyebabkan komentar sulit terdeteksi. Kondisi ini berpotensi menyebabkan lingkungan digital yang tidak aman khususnya yang menjadi objek seksualisasi. Penelitian ini bertujuan untuk membangun model klasifikasi yang mampu mendeteksi komentar bermuatan pelecehan seksual verbal dengan membandingkan dua algoritma klasifikasi, yaitu *K-Nearest Neighbor* (KNN) dan *Support Vector Machine* (SVM). Data penelitian berjumlah 7.063 yang didapat dari komentar konten publik figur Korea di Tik Tok, kemudian diberi label secara manual ke dalam kelas pelecehan seksual dan non-pelecehan seksual. Data komentar diproses melalui tahap pembersihan teks, transformasi data teks ke dalam bentuk vektor numerik menggunakan *word embedding* LaBSE (*Language-agnostic BERT Sentence Embedding*), serta pembagian data. Kedua algoritma dilatih dengan data yang sama dan dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengetahui performa model dalam mendeteksi komentar pelecehan seksual verbal. Hasil penelitian menunjukkan algoritma SVM dengan skenario pembagian data 80:20 memberikan performa lebih baik dan stabil dibandingkan KNN. Nilai akurasi mencapai 84% dengan nilai *recall* dan *F1-score* kelas pelecehan seksual lebih tinggi dibanding KNN, menunjukkan kemampuan SVM lebih baik dalam mendeteksi kelas minoritas. Secara keseluruhan, penelitian menunjukkan kombinasi algoritma SVM dan LaBSE lebih efektif dalam mendeteksi komentar pelecehan seksual verbal berbasis teks dan dapat dimanfaatkan sebagai dasar pengembangan sistem moderasi konten platform media sosial guna menciptakan ruang digital yang lebih aman.

Kata kunci: *Support Vector Machine, K-Nearest Neighbor, LaBSE, Pelecehan Seksual Verbal, Machine Learning*

ABSTRACT

Verbal sexual harassment is a common form of violence on social media, including in TikTok's user comment sections. The sheer volume of comments, short, informal, and often implicit makes them hard to detect. This can create unsafe digital spaces, especially for those targeted by sexualization. This study develops a classification model to identify verbal sexual harassment comments by comparing K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms. The dataset includes 7,063 comments from TikTok videos of Korean public figures, manually labeled as sexual harassment or non-harassment. Data processing involved text cleaning, conversion to numerical vectors using LaBSE (Language-agnostic BERT Sentence Embedding), and splitting. Both algorithms were trained on the same data and evaluated with accuracy, precision, recall, and F1-score metrics to gauge performance in detecting harassment comments. Results show SVM with an 80:20 split outperforms and stabilizes better than KNN, achieving 84% accuracy. It delivered higher recall and F1-score for the sexual harassment minority class, proving superior at identifying rare cases. Overall, combining SVM and LaBSE effectively detects text-based verbal harassment, providing a foundation for social media content moderation to foster safer online environment

Keyword: Support Vector Machine, K-Nearest Neighbor, LaBSE, Verbal Sexual Harassment, Machine Learning