

**ANALISIS TINGKAT CYBERBULLYING DI TWITTER (X)
MENGUNAKAN ALGORITMA SVM DAN NAIVE BAYES**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

RESTU ADJI

22.21.1556

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

**ANALISIS TINGKAT CYBERBULLYNG DI TWITTER (X)
MENGUNAKAN ALGORITMA SVM DAN NAIVE BAYES**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
RESTU ADJI
22.21.1556

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024**

HALAMAN PERSETUJUAN

SKRIPSI

ANALISIS TINGKAT CYBERBULLYNG DI TWITTER (X) MENGGUNAKAN ALGORITMA SVM DAN NAIVE BAYES

yang disusun dan diajukan oleh

Restu Adji

22.21.1556

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Juli 2024

Dosen Pembimbing,



Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.

NIK. 190302235

HALAMAN PENGESAHAN

SKRIPSI

ANALISIS TINGKAT CYBERBULLYING DI TWITTER (X) MENGGUNAKAN ALGORITMA SVM DAN NAÏVE BAYES

yang disusun dan diajukan oleh

Restu Adji

22.21.1556

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arifvanto Hadinegoro, S.kom, MT
NIK. 190302289

M. Rudyanto Arief, S.T, M.T
NIK. 190302098

Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.
NIK. 190302235

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Juli 2024

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Restu Adji
NIM : 22.21.1556

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Tingkat Cyberbullyng Di Twitter (X) Menggunakan Algoritma SVM Dan NAIVE BAYES

Dosen Pembimbing: Dr. Ferry Wahyu Wibowo, S.Si., M.Cs.

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 Juli 2024

Yang Menyatakan,



Restu Adji

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam atas rezeki yang melimpah yang diberikan Allah Swt sehingga dapat menyelesaikan skripsi ini. Penulis mempersembahkan skripsi ini kepada:

1. Keluarga besar dan orang tua penulis yang telah senantiasa memberikan doa, dukungan serta motivasi baik secara moral maupun materil selama menempuh perkuliahan ini hingga selesai.
2. Bapak Dr. Ferry Wahyu Wibowo, S.Si., M.Cs., yang telah menjadi dosen pembimbing saya, yang selalu membantu saya selama penulisan skripsi dan memungkinkan untuk dapat menyelesaikan pekerjaan skripsi ini.
3. Teman-teman penulis baik itu seangkatan dan sesama mahasiswa transfer dari Universitas AMIKOM Yogyakarta maupun teman-teman online (discord) yang telah memberikan semangat, masukan, dan arahan hingga akhirnya dapat terselesaikan skripsi ini.

KATA PENGANTAR

Pertama perkenankan saya untuk mengucapkan “Alhamdulillah” kehadirat Allah.SWT. yang telah memberikan Kekuatan serta Kesabaran, sehingga Skripsi dengan judul “**Analisis Tingkat Cyberbullyng Di Twitter (X) Menggunakan Algoritma SVM Dan NAIVE BAYES**” ini dapat terselesaikan.

Skripsi ini adalah salah satu persyaratan dalam menyelesaikan jenjang pendidikan Sarjana I, program studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Ucapan terima kasih disampaikan kepada semua pihak yang telah memberikan dukungan dan bantuan atas terselesaikannya Skripsi ini. Ucapan terima kasih khusus disampaikan kepada :

1. Allah.SWT.
2. Kedua Orang Tua dan Keluarga Besar Penulis Yang Telah Mendukung Penuh, Memotivasi, Serta Materi.
3. Hanif Al Fatta,S.Kom., M.Kom., Ph.D. Selaku Ketua Dekan Fakultas Ilmu Komputer.
4. Ibu Windha Mega Pradnya Duhita, M.Kom. Selaku Ketua Prodi SI Informatika.
5. Dr. Ferry Wahyu Wibowo, S.Si., M.Cs. Selaku Dosen Pembimbing Penulis.
6. Arifiyanto Hadinegoro, S.Kom, MT Selaku Dosen Penguji 1.
7. M. Rudyanto Arief, S.T, M.T Selaku Dosen Penguji 2.
8. Serta Teman-Teman Penulis yang selalu memberi semangat, motivasi, dan dukungan penuh maupun teman seangkatan mahasiswa transfer 2022 dan Teman online (Discord).

Yogyakarta, 24 Juli 2024
Penulis



DAFTAR ISI

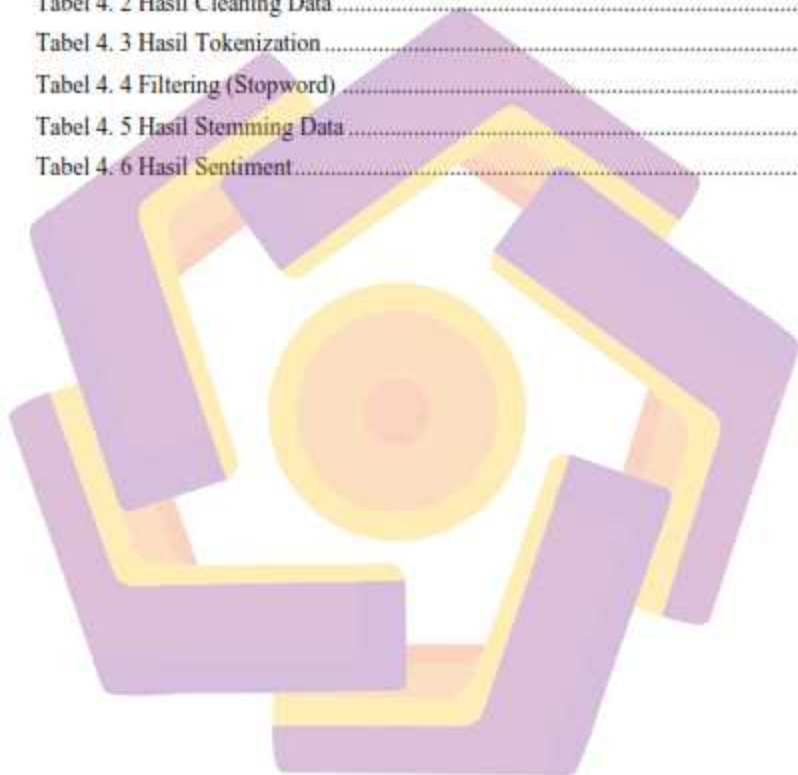
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Study Literatur.....	5

2.2	Dasar Teori.....	10
2.2.1	Machine Learning	10
2.2.2	Supervised Learning	11
2.2.3	Analisis Setimen	12
2.2.4	Cyberbullying	12
2.2.5	Twitter (X)	13
2.2.6	Data Crawling	14
2.2.7	Python	15
2.2.8	<i>Support Vector Machine (SVM)</i>	16
2.2.9	Naïve Bayes	19
BAB III METODE PENELITIAN		21
3.1	Objek Penelitian	21
3.2	Alur Penelitian	21
3.3	Alat dan Bahan	24
BAB IV HASIL DAN PEMBAHASAN		25
4.1	Crawling & Pengumpulan Data	25
4.2	Pre-Processing Data	27
4.2.1	Drop Kolom	27
4.2.2	Identifikasi & Mengisi Data Hilang	27
4.2.3	Cleaning Data.....	29
4.2.4	Tokenization	30
4.2.5	Filtering (Stopword).....	32
4.2.6	Stemming Data.....	33
4.2.7	Save Data Preprocessing.....	35
4.2.8	Membuang Data Duplikat.....	35

4.3	Visualisasi Data Clean	35
4.3.1	Memastikan Dataset Bahasa Indonesia.....	36
4.4	Pelabelan Data.....	36
4.5	Split Data.....	38
4.6	TF-IDF	39
4.7	SMOTE	39
4.6	Klasifikasi SVM.....	39
4.7	Klasifikasi Naïve Bayes	41
4.8	ROC (Receiver Operating Characteristic).....	43
BAB V	PENUTUP.....	45
5.1	Kesimpulan.....	45
5.2	Saran.....	46
REFERENSI.....		47
LAMPIRAN.....		51

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian	6
Tabel 3. 1 Kunci Pencarian	21
Tabel 4. 1 Hasil Crawling & Pengumpulan Data	25
Tabel 4. 2 Hasil Cleaning Data	29
Tabel 4. 3 Hasil Tokenization	31
Tabel 4. 4 Filtering (Stopword)	32
Tabel 4. 5 Hasil Stemming Data	34
Tabel 4. 6 Hasil Sentiment	37

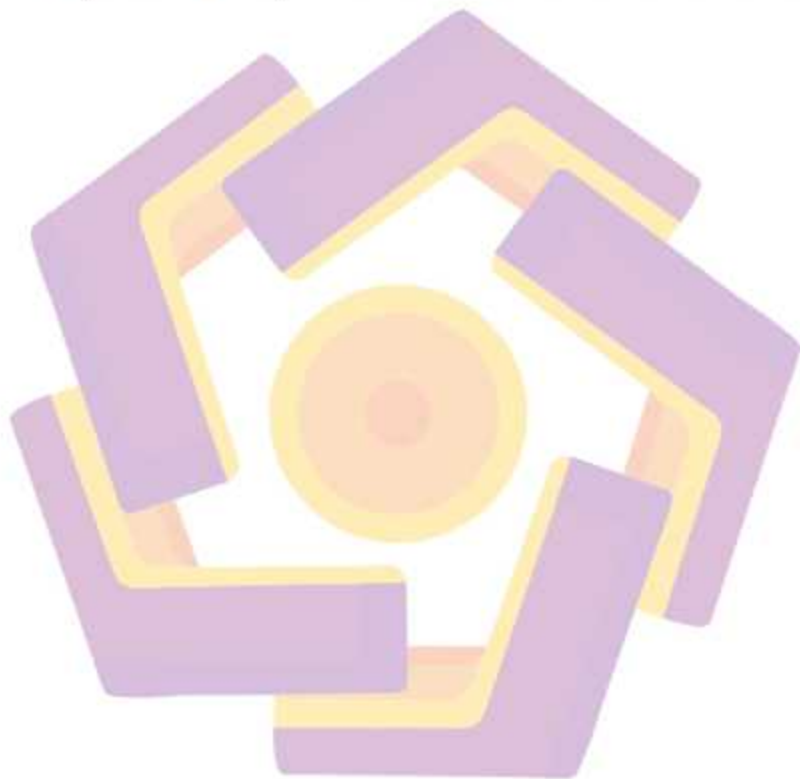


DAFTAR GAMBAR

Gambar 2. 1 Komponen dasar Machine Learning	10
Gambar 2. 2 Konsep Kerja Supervised Learning	11
Gambar 2. 3 Contoh Sentiment Analysis	12
Gambar 2. 4 Support Vector Machine (SVM).....	17
Gambar 2. 5 Linear SVM.....	18
Gambar 2. 6 Non-Linear SVM	19
Gambar 3. 1 Alur Penelitian	22
Gambar 4. 1 Flowchart Alur Preprocessing Data	27
Gambar 4. 2 Mengecek Data Missing.....	28
Gambar 4. 3 Hasil setelah diisi dengan data yang sering muncul.....	28
Gambar 4. 4 Wordcloud.....	35
Gambar 4. 5 Frekuensi Kata	36
Gambar 4. 6 Visual Hasil Sentiment.....	38
Gambar 4. 7 Split data SMOTE 7:3	38
Gambar 4. 8 Hasil SMOTE.....	39
Gambar 4. 9 Penerapan Model SVM.....	40
Gambar 4. 10 Hasil Akurasi SVM & Confusion Matrix	40
Gambar 4. 11 Visual Confusion Matrix.....	41
Gambar 4. 12 Penerapan Model NB Multinomial	42
Gambar 4. 13 Hasil Akurasi NB & Confusion Matrix	42
Gambar 4. 14 Visual Confusion Matrix NB	43
Gambar 4. 15 ROC Curve SVM dan Naive Bayes	44

DAFTAR LAMPIRAN

Lampiran 1 link code crawling data penulis.....	51
Lampiran 2 link code Project Analisis Sentimen Cyberbullying penulis.....	51
Lampiran 3 link dataset penulis.....	51



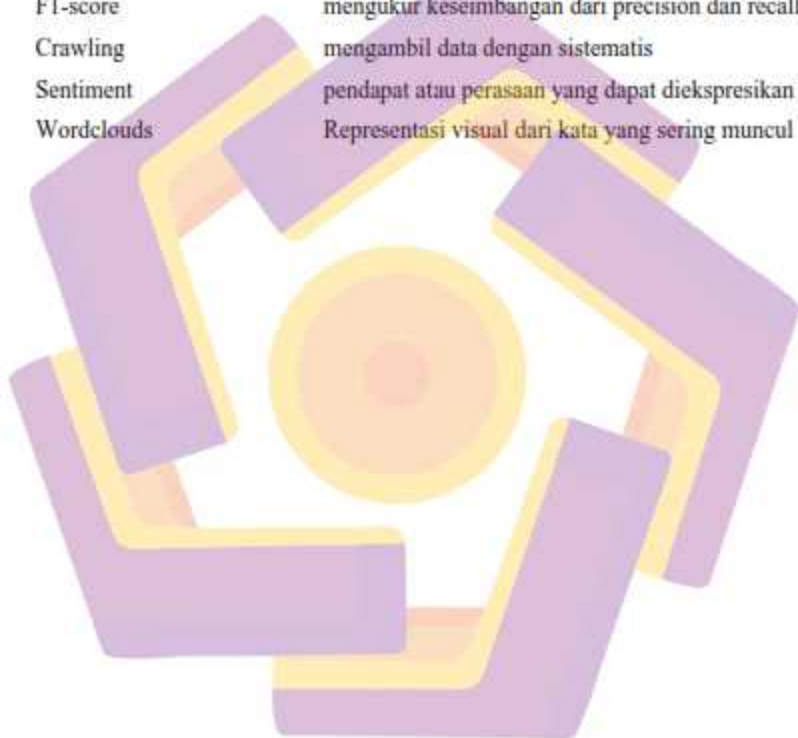
DAFTAR LAMBANG DAN SINGKATAN



NB	Naïve Bayes
SVM	Support Vector Machines
ML	Learning
pd	Pandas
df	dataframe
np	numpy
sns	seaborn
plt	matplotlib.pyplot
ROC	Receiver Operating Characteristic
NLTK	Natural Language Toolkit
cm	confusion matrix
TF-IDF	Term Frequency-Inverse Document Frequency
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positif
TN	True Negatif
FP	False Positif
FN	False Negatif

DAFTAR ISTILAH

Confusion Matrix	matriks yang Berisi TP, TN, FP, FN
Accuracy	Mengukur akurasi model dengan benar
Precision	Mengukur presisi model dengan benar
Recall	Mengukur sentivitas atau true positif rate
F1-score	mengukur keseimbangan dari precision dan recall
Crawling	mengambil data dengan sistematis
Sentiment	pendapat atau perasaan yang dapat diekspresikan
Wordclouds	Representasi visual dari kata yang sering muncul



INTISARI

Seiring Perkembangan Teknologi, Pengguna Sosial Media Juga Bertambah Terutama Di sosial media X (Twitter) tentu juga semakin beragam komentar dan postingan yang di tulis oleh seseorang. Terkadang Tanpa disadari para pengguna lain di Twitter menuliskan komentar atau postingan yang mengandung unsur bullying, dimana terkadang menyakiti perasaan si Pembaca Komentar atau postingan tersebut yang akan terjadi ketika pengguna menggunakan platform tersebut untuk menyebarkan pesan yang merendahkan, menghina, atau mengintimidasi orang lain secara anonim atau terbuka. Hal ini dapat berdampak negatif secara psikologis, emosional, dan bahkan dapat memicu konsekuensi yang berbahaya seperti depresi dan kecemasan pada korban.

Oleh Karena itu perlu Analisis Tingkat Cyberbullying dan mengklasifikasi apakah isi komentar atau postingan tersebut mengandung unsur cyberbullying atau tidak, dimana Penelitian ini menggunakan 2 metode yaitu *SVM (Support Vector Machine)* dan *Naive Bayes*.

Penelitian ini bertujuan untuk mengembangkan model yang efektif dalam mendeteksi pesan-pesan yang bersifat *cyberbullying* di *Twitter*. *SVM* digunakan untuk menghasilkan model klasifikasi yang memisahkan pesan-pesan *cyberbullying* dari yang bukan, dengan memanfaatkan teknik pemisahan linier pada ruang fitur yang dihasilkan dari teks pesan. Di sisi lain, *Naive Bayes* menggunakan asumsi independensi fitur untuk memprediksi kemungkinan pesan tertentu sebagai *cyberbullying* atau bukan.

Kata kunci: *Twitter, Analisis, Cyberbullying, SVM, Naive Bayes.*

ABSTRACT

As technology develops, social media users also increase, especially on social media. Sometimes, without realizing it, other users on Twitter write comments or posts that contain bullying elements, which sometimes hurt the readers' feelings. These comments or posts will occur when users use the platform to spread messages that demean, insult or intimidate other people anonymously or open. This can have a negative impact psychologically, emotionally, and can even trigger dangerous consequences such as depression and anxiety in the victim.

Therefore, it is necessary to analyze the level of cyberbullying and classify whether the contents of the comments or posts contain elements of cyberbullying or not, where this research uses 2 methods, namely SVM (Support Vector Machine) and Naive Bayes.

This research aims to develop an effective model in detecting cyberbullying messages on Twitter. SVM is used to produce a classification model that separates cyberbullying messages from those that are not, by utilizing linear separation techniques in the feature space generated from the message text. On the other hand, Naive Bayes uses the assumption of feature independence to predict the likelihood of a particular message being cyberbullying or not.

Keyword: Twitter, analysis, Cyberbullying, SVM, Naive Bayes