

**KLASIFIKASI TWEET BULLYING MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DAMAI PATRIA SUYANA

19.11.2602

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2023**

**KLASIFIKASI TWEET BULLYING MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

DAMAI PATRIA SUYANA

19.11.2602

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI TWEET BULLYING MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

yang disusun dan diajukan oleh

Damal Patria Suyana

19.11.2601

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Juni 2023

Dosen Pembimbing,



Aini Yaqin, M.Kom.

NIK. 190302255

HALAMAN PENGESAHAN
SKRIPSI
KLASIFIKASI TWEET BULLYING MENGGUNAKAN
METODE K-NEAREST NEIGHBOR

yang disusun dan diajukan oleh

Damal Patria Suyana

19.11.2601

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Juni 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Arif Dwi Laksito, M.Kom.
NIK. 190302150

Ika Nur Fajri, M.Kom.
NIK. 190302268

Alnul Yaqin, M.Kom.
NIK. 190302255



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 21 Juni 2023

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Damai Patria Syamsi
NIM : 149.11.2082

Meyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Twerk Bullying Menggunakan Metode K-Nearest Neighbor
Dosen Pembimbing : Aini Yogi, M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERAH 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 bentuk dengan diikutkan nama pengarang dan diikutkan 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 penarikan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan aturan yang berlaku di Perguruan Tinggi).

Yogyakarta, 21 Juni 2023

Yang Memastikan,



Damai Patria Syamsi

KATA PENGANTAR

Puji syukur atas berkat, rahmat dan anugrah yang berikan oleh Tuhan Yang Maha Esa pada penulis, sehingga dapat terselesaikannya skripsi yang berjudul “Klasifikasi Tweet Bullying Menggunakan Metode K-Nearest Neighbor” dengan baik dan tanpa kendala yang berarti, sebagai salah satu syarat untuk menuntaskan Program Studi Sarjana Informatika pada Universitas Amikom Yogyakarta.

Selama proses pengerjaan skripsi penulis sadar bahwa sudah banyak pihak yang memberikan bimbingan, bantuan, semangat, dan dukungan sehingga skripsi ini dapat terselesaikan. Bukan tidak mungkin tanpa adanya banyak pihak tersebut skripsi ini jadi tidak terselesaikan, oleh karena itu pada kesempatan ini penulis ingin mengucapkan banyak terimakasih kepada:

1. Ainul Yaqin, S.Kom., M.Kom, dosen pembimbing penulis yang selalu memberikan bimbingan serta arahan untuk skripsi ini pada penulis.
2. Orang Tua penulis yang selalu memberikan doa sehingga skripsi ini bisa terselesaikan serta memberikan dukungan materi maupun moral pada penulis selama kuliah hingga skripsi.
3. Keluarga penulis yang selalu memberikan dukungan dan semangat serta doa pada penulis.
4. Kekasih hati penulis yang selalu memberikan semangat dan doa serta motivasi untuk terus menjejarkan skripsi ini disaat rasa malas muncul.

Yogyakarta, 23 November 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	viii
DAFTAR GAMBAR.....	ix
INTISARI.....	x
ABSTRACT.....	xi
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 Studi Literatur.....	5
2.2 Dasar Teori.....	11
BAB III METODE PENELITIAN.....	14
3.1 Objek Penelitian.....	14
3.2 Alur Penelitian.....	15

3.3	Alat dan Bahan	23
Bab IV HASIL DAN PEMBAHASAN		25
4.1	Pengumpulan Data	25
4.2	Pelabelan Data	26
4.3	Data Preprocessing	27
4.4	TF-IDF	29
4.5	K-Nearest Neighbor	30
4.6	Confusion Matrix	30
4.7	Pengujian Nilai k dan Perbandingan Data Latih dengan Data Uji	31
4.8	Pengujian Keseimbangan Data	37
BAB V PENUTUP		40
5.1	Kesimpulan	40
5.2	Saran	40
REFERENSI		42

DAFTAR TABEL

Tabel 3.1 Contoh Data	18
Tabel 3.2 Perhitungan TF-IDF.....	19
Tabel 3.3 Hasil TF-IDF.....	20
Tabel 3.4 Perhitungan Jarak Euclidean.....	21
Tabel 3.5 Hasil Klasifikasi.....	21
Tabel 3.6 Contoh Hasil Klasifikasi.....	22
Tabel 3.7 Contoh Confusion Matrix.....	22
Tabel 4.1 Hasil Skenario Pertama.....	31
Tabel 4.2 Hasil Skenario Kedua.....	32
Tabel 4.3 Hasil Skenario Ketiga.....	33
Tabel 4.4 Hasil Skenario Keempat.....	34
Tabel 4.5 Hasil Skenario Kelima.....	35
Tabel 4.6 Hasil Terbaik Semua Skenario.....	36
Tabel 4.7 Pengujian Kedua Skenario Pertama.....	37
Tabel 4.8 Pengujian Kedua Skenario Kedua.....	39

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	15
Gambar 3.2 Tahap Cleansing	16
Gambar 3.3 Case Folding	17
Gambar 3.4 Tokenisasi	17
Gambar 3.5 Stopword Removal	18
Gambar 3.6 Stemming	18
Gambar 4.1 Twitter API Key	25
Gambar 4.2 Kode Program Ambil Data	25
Gambar 4.3 Data Tweets	26
Gambar 4.4 Data Setelah Pelabelan	26
Gambar 4.5 Kode Program Cleansing	27
Gambar 4.6 Kode Program Case Folding	27
Gambar 4.7 Kode Program Tokenisasi	28
Gambar 4.8 Kode Program Stopword Removal	28
Gambar 4.9 Kode Program Stemming	28
Gambar 4.10 Data Setelah Preprocessing	29
Gambar 4.11 Kode Program TF-IDF	29
Gambar 4.12 Kode Program KNN	30
Gambar 4.13 Kode Program Confusion Matrix	30

INTISARI

Memiliki akun media sosial pada jaman sekarang merupakan hal yang wajar atau bahkan wajib. Oleh hal tersebutlah kemajuan media sosial menjadi sangat pesat dan membuat masyarakat tidak tahu batasan - batasan dalam bermedia sosial. Salah satu media sosial yang dipakai masyarakat Indonesia adalah Twitter. Pada media sosial Twitter kerap kali terjadi tindak kejahatan yang bisa merugikan. Salah satu dari tindak kejahatan adalah perundungan siber atau cyber bullying yang bisa mengakibatkan korban depresi hingga bunuh diri.

Meskipun merupakan tindak kejahatan, tetapi perundungan siber masih kerap terjadi. Oleh hal tersebut maka perlu adanya tindakan deteksi yang mampu mencegah terjadinya perundungan siber. Untuk mendeteksi tindakan perundungan siber maka perlu dilakukan text mining pada tweets dan dilanjutkan dengan mengklasifikasikan tweets tersebut menggunakan metode KNN. KNN atau K-Nearest Neighbor merupakan metode klasifikasi dengan cara kerja mencari jarak terdekat dari suatu objek dengan data latih yang digunakan untuk melatih KNN itu sendiri.

Pada Klasifikasi dengan menggunakan metode KNN ini data latih akan dilabeli dengan dua class yaitu "perundungan" dan "bukan perundungan". Klasifikasi yang dilakukan dengan menggunakan metode K-Nearest Neighbor akan menghasilkan sebuah model klasifikasi yang nantinya dapat di implemetasi dalam berbagai hal, contohnya di implementasikan dalam website.

Kata kunci: Text Mining, Teknologi, Twitter, Klasifikasi

ABSTRACT

Having a media social account nowadays is normal thing or even mandatory. Because of that the evolution of social media has become very fast and has made people not know the boundaries of social media. One of the social media used by Indonesian people is Twitter. On Twitter there are often crimes that can be detrimental. One of the crimes is cyber bullying or perundungan siber which can lead to depression and suicide.

Even though it's a crime, cyber bullying is still happening. Because of that, it's necessary to have detection action that can prevent cyber bullying. To detect cyber bullying, it is necessary to do text mining on tweets and then classify the tweets using the KNN method. KNN or K-Nearest Neighbor is classification method, by working to find the closest distance from an object between the training data used to train KNN.

In the classification using the KNN method, the training data will be labeled with two classes, namely "bullying" and "non-bullying". Classification using the K-Nearest Neighbor method will produce a classification model that can be implemented in various things, for example implemented on a website.

Keyword: *Text Mining, Tech, Twitter, Classification*