

**DETEKSI *CYBERBULLYING* PADA KOMENTAR SOSIAL MEDIA
DENGAN KLASIFIKASI TEKS**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

MUHAMMAD ILHAM PRADANTO

19.83.0421

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2023

**DETEKSI *CYBERBULLYING* PADA KOMENTAR SOSIAL MEDIA
DENGAN KLASIFIKASI TEKS**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Teknik Komputer



disusun oleh

MUHAMMAD ILHAM PRADANTO

19.83.0421

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA**

YOGYAKARTA

2023

HALAMAN PERSETUJUAN

SKRIPSI

**DETEKSI *CYBERBULLYING* PADA KOMENTAR SOSIAL MEDIA
DENGAN KLASIFIKASI TEKS**

yang disusun dan diajukan oleh

Muhammad Ilham Pradanto

19.83.0421

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 10 Agustus 2023

Dosen Pembimbing,



Anggit Ferdita Nugraha, S.T., M.Eng.

NIK. 190302480

HALAMAN PENGESAHAN
SKRIPSI
DETEKSI *CYBERBULLYING* PADA KOMENTAR SOSIAL MEDIA
DENGAN KLASIFIKASI TEKS

yang disusun dan diajukan oleh

Muhammad Ilham Pradanto

19.83.0421

Telah dipertahankan di depan Dewan Penguji
pada tanggal 10 Agustus 2023

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Banu Santoso, S.T., M.Eng.
NIK. 190302327

Ika Nur Fajri, M.Kom.
NIK. 190302268

Anggit Ferdita Nugraha, S.T., M.Eng.
NIK. 190302480

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

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Ilham Pradanto
NIM : 19.83.0421

Menyatakan bahwa Skripsi dengan judul berikut:

Deteksi *Cyberbullying* Pada Komentar Sosial Media Dengan Klasifikasi Teks

Dosen Pembimbing : Anggit Ferdita Nugraha, S.T., M.Eng.

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, 21 Agustus 2023

Yang Menyatakan,



Muhammad Ilham Pradanto

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya kepada penulis, sehingga dapat menyelesaikan skripsi dengan judul “Deteksi *Cyberbullying* pada Komentar Sosial Media dengan Klasifikasi Teks” dengan baik. Skripsi ini disusun dan diselesaikan sebagai syarat dalam memperoleh gelar sarjana pada Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

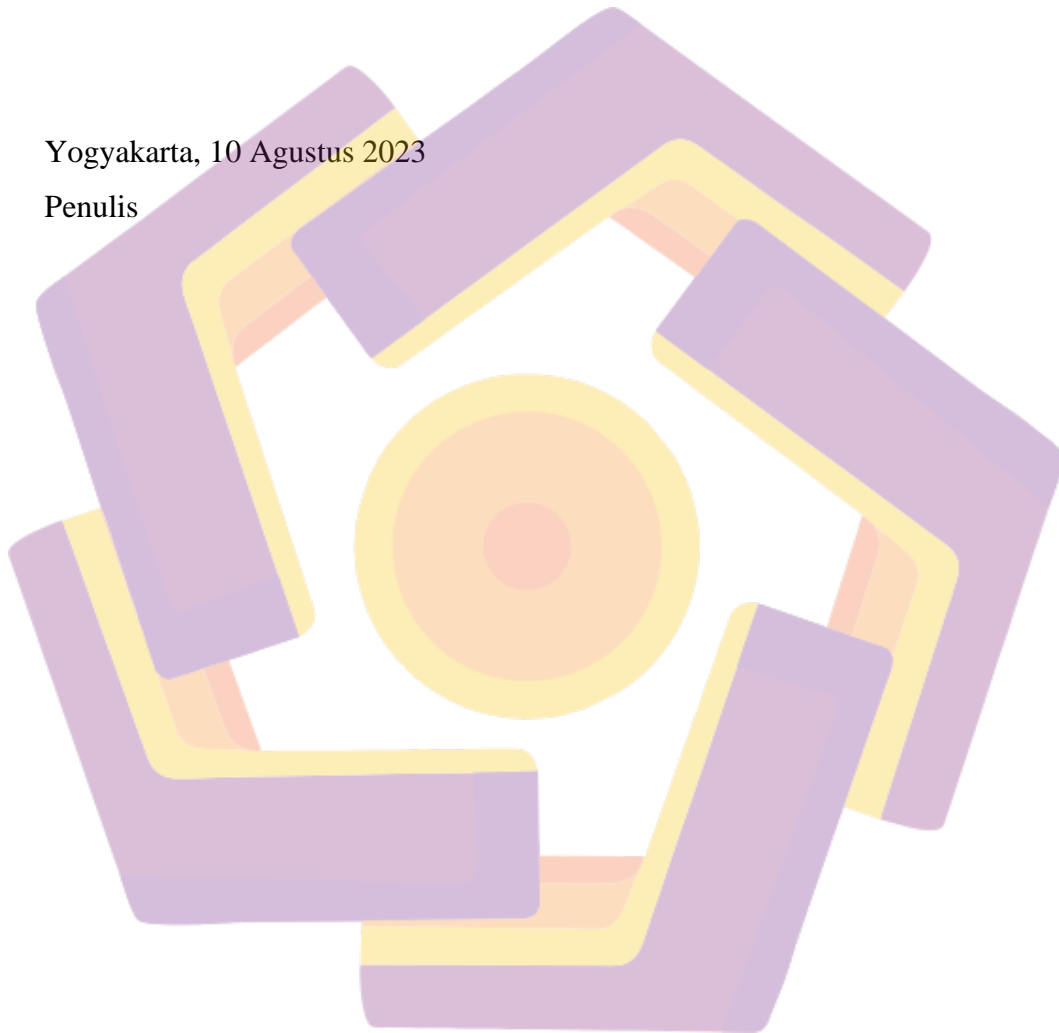
Dalam proses menyelesaikan skripsi, penulis telah mendapat banyak bantuan maupun dukungan baik secara moral maupun materiil dari berbagai pihak. Oleh karena itu penulis mengucapkan banyak terima kasih kepada:

1. Kedua orangtua serta adik penulis yang telah memberikan doa maupun dukungan kepada penulis dalam membantu kelancaran pengerjaan skripsi penulis.
2. Bapak Anggit Ferdita Nugraha, S.T., M.Eng., selaku Dosen Pembimbing yang telah memberikan bimbingan, arahan, ilmu, dan masukan dalam menyelesaikan skripsi ini.
3. Dosen penguji sidang skripsi yang telah memberikan arahan dan saran kepada penulis agar skripsi menjadi baik.
4. Seluruh bapak dan ibu dosen yang telah mendidik dan memberikan ilmu selama penulis menempuh Pendidikan di Universitas AMIKOM Yogyakarta.
5. Teman-teman Program Studi Teknik Komputer yang telah memberikan dukungan dalam penelitian ini.
6. Serta seluruh pihak yang tidak dapat disebutkan oleh penulis satu per satu yang telah membantu kelancaran pengerjaan skripsi.

Penulis menyadari bahwa skripsi ini tidak lepas dari kekurangan baik dalam format maupun isi. Oleh karena itu diharapkan kritik maupun saran yang membangun dalam proses untuk memperbaiki diri. Penulis berharap semoga skripsi ini dapat bermanfaat bagi pembaca dan memberikan kontribusi positif dalam bidang studi yang diteliti.

Yogyakarta, 10 Agustus 2023

Penulis



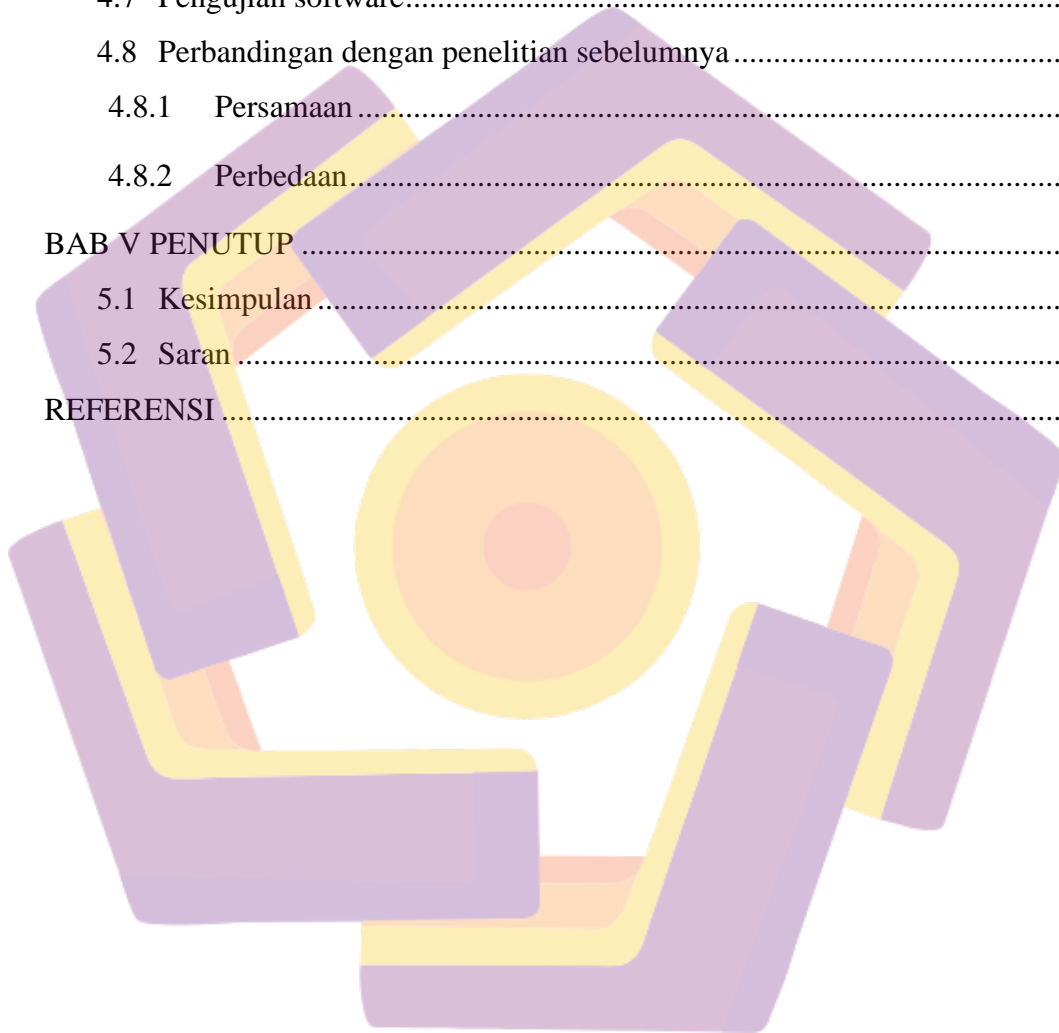
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR PERSAMAAN.....	xiii
INTISARI	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penulisan	3
BAB II TINJAUAN PUSTAKA	4
2.1 Studi Literatur	4
2.2 Dasar Teori	10
2.2.1 Media Sosial	10
2.2.2 <i>Cyberbullying</i>	10
2.2.3 Machine learning	12
2.2.4 Analisis Sentimen	13
2.2.5 Natural Language Processing	13
2.2.6 Text Mining	14

2.2.7	Klasifikasi	14
2.2.8	Ekstraksi Fitur.....	15
2.2.9	Pembobotan TF-IDF	15
2.2.10	Logistic Regression	16
2.2.11	Decision Tree.....	17
2.2.12	Random Forest Classifier	18
2.2.13	K-Nearest Neighbors	19
2.2.14	Extreme Gradient Boosting	19
2.2.15	Multinomial Naive Bayes.....	21
2.2.16	Confussion Matrix	22
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	24
3.2.1	Akuisisi Data	24
3.2.1	Data Preparation dan Data Preprocessing	25
3.2.1.1	Alur Data Prepatarion	26
3.2.1.2	Alur Data Preprocessing	28
3.2.2	Featureing dan Modeling.....	30
3.2.2.1	Alur Featureing	31
3.2.2.2	Alur Modeling.....	32
3.2.2.3	Alur Evaluasi	33
3.2.3	Testing	34
3.2.4	Implementasi.....	35
3.3	Alat dan Bahan.....	36
3.3.1	Hardware	36
3.3.2	Software	37

3.3.3	Dataset.....	37
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Data Preparation	38
4.1.1	Seleksi Data	38
4.1.2	Pemeriksaan Data	39
4.2	Data Preprocessing	39
4.2.1	Case Folding	40
4.2.2	Data Cleanning	40
4.2.3	Tokenisasi	41
4.2.4	Stopword Removal	42
4.2.5	Normalisasi	42
4.2.6	Stemming.....	43
4.3	Featureing	44
4.3.1	Menentukan Scikit-Learn TF-IDF Formula	44
4.3.2	Menentukan fitur pembobotan kata	45
4.3.3	Menentukan TF-IDF unigram.....	45
4.4	Modeling.....	46
4.4.1	Menentukan banyak perulangan	46
4.4.2	Split data	47
4.4.3	Pemilihan model	47
4.4.4	Pelatihan model	47
4.4.5	Evaluasi.....	48
4.5	Testing	49
4.5.1	Proses test data	49
4.6	Implementasi.....	51
4.6.1	Pemilihan Framework Web	51

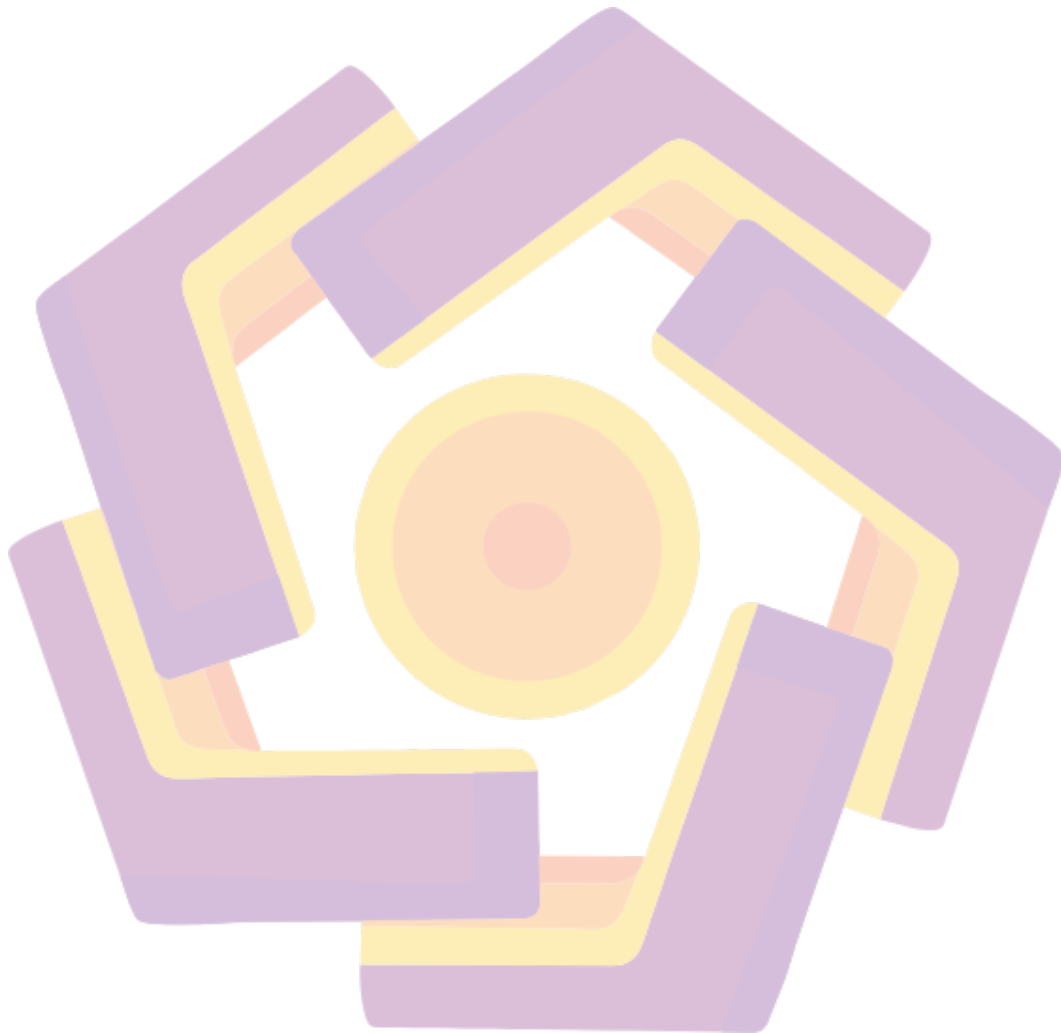
4.6.2	Pengintegrasian Model	51
4.6.3	Pembuatan Antarmuka Pengguna.....	51
4.6.4	Prediksi dan Visualisasi Hasil	52
4.6.5	Error handling	52
4.7	Pengujian software.....	52
4.8	Perbandingan dengan penelitian sebelumnya	57
4.8.1	Persamaan	57
4.8.2	Perbedaan.....	57
BAB V PENUTUP		58
5.1	Kesimpulan.....	58
5.2	Saran	59
REFERENSI		60



DAFTAR TABEL

Tabel 2.1. Keaslian Penelitian
Tabel 2.2. Confusion Matrix

7
22

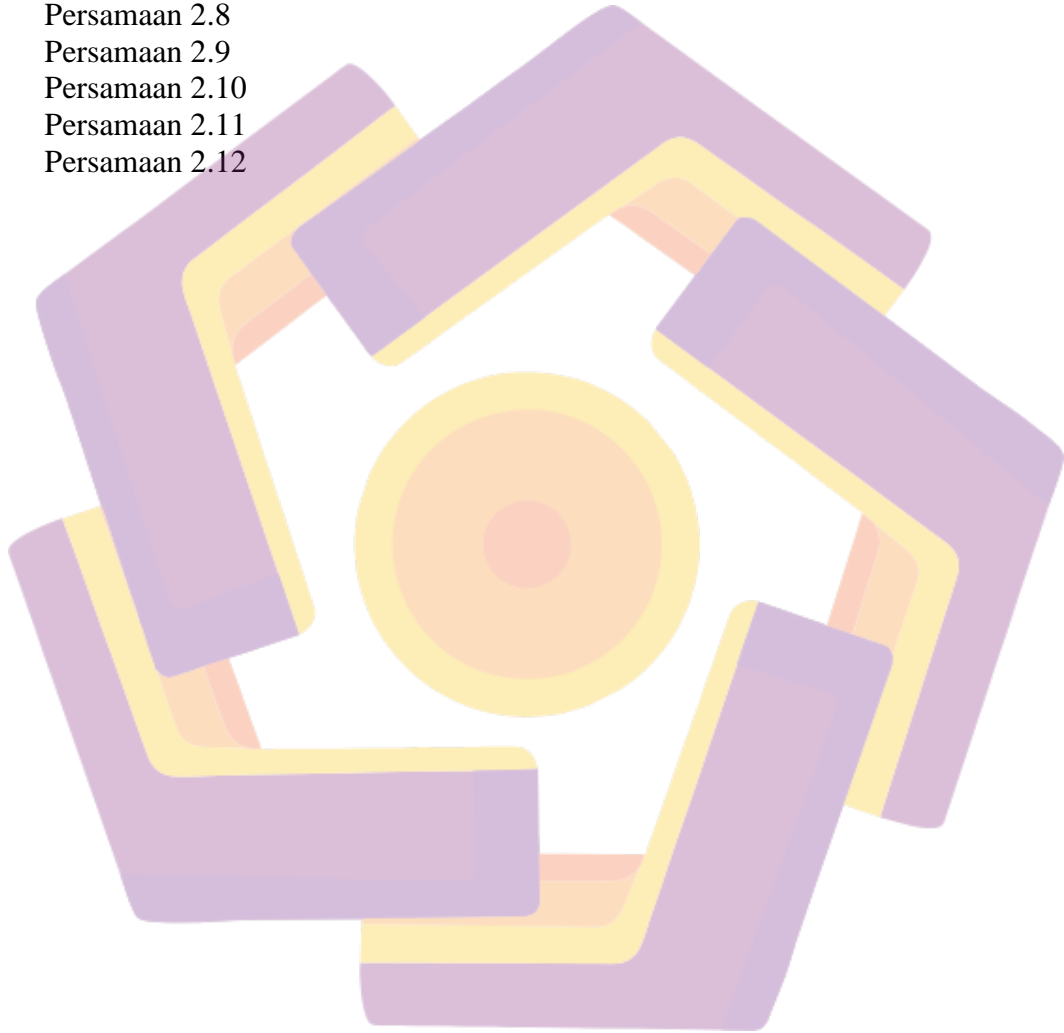


DAFTAR GAMBAR

Gambar 2.1. Konsep Dasar Decision Tree	18
Gambar 2.2. Konsep Dasar Forest Classifier	19
Gambar 3.1. Data Preparation dan Data Preprocessing	25
Gambar 3.2. Data Preparation	26
Gambar 3.3. Data Preprocessing	28
Gambar 3.4. Featureing dan Modeling	30
Gambar 3.5. Featureing	31
Gambar 3.6. Modeling	32
Gambar 3.7. Evaluasi	33
Gambar 3.8. Testing	34
Gambar 3.9. Implementasi	36
Gambar 4.1. View Dataset	38
Gambar 4.2. Atribut yang Digunakan	39
Gambar 4.3. Distribusi Komentar	40
Gambar 4.4. Dataset sebelum Preprocessing	40
Gambar 4.5. Case Folding Data	41
Gambar 4.6. Data Cleaning	41
Gambar 4.7. Tokenisasi	42
Gambar 4.8. Stopword Removal	43
Gambar 4.9. Normalisasi	43
Gambar 4.10. Stemming	44
Gambar 4.11. Dataset hasil Preprocessing	45
Gambar 4.12. Dataset hasil Featureing	47
Gambar 4.13. Rata-rata Akurasi Model dari 30 Iterasi	49
Gambar 4.14. Hasil Evaluasi Model	50
Gambar 4.15. Data Testing	51
Gambar 4.16. Hasil Prediksi dari Data Testing	51
Gambar 4.17. Halaman Awal Website	53
Gambar 4.18. Error Handling	54
Gambar 4.19. Halaman Hasil Prediksi	54
Gambar 4.20. Grafik Kinerja Model	55
Gambar 4.21. Piechart Distribusi Prediksi	55
Gambar 4.22. Hasil Prediksi dari tiap Model	56
Gambar 4.23. Detail Hasil Prediksi dari tiap Model	56
Gambar 4.24. Total Hasil Confidence Score Setiap Kategori	57
Gambar 4.25. Hasil Preprocessing Input Data	57
Gambar 4.26. Kesimpulan dari Prediksi	57

DAFTAR PERSAMAAN

Persamaan 2.1	16
Persamaan 2.2	16
Persamaan 2.3	17
Persamaan 2.4	19
Persamaan 2.5	20
Persamaan 2.6	20
Persamaan 2.7	21
Persamaan 2.8	21
Persamaan 2.9	22
Persamaan 2.10	22
Persamaan 2.11	23
Persamaan 2.12	23



INTISARI

Penggunaan media internet di Indonesia terus mengalami peningkatan setiap tahunnya pada tahun 2023 tercatat sebanyak 215 juta penduduk menggunakan internet. Seiring dengan meningkatnya penggunaan media internet beberapa platform media sosial seperti Facebook, Twitter, Instagram juga semakin populer di kalangan masyarakat. Namun, media sosial ini tidak lepas dari bahaya *cyberbullying* yang sering dilakukan oleh pengguna khususnya pada kolom komentar. Bahaya *cyberbullying* tentunya meresahkan banyak orang dikarenakan dampak yang ditimbulkan, maka salah satu upaya untuk mencegah terjadinya *bullying* di media sosial adalah dengan melakukan analisis sentimen pada kolom komentar media sosial yang bertujuan untuk mengetahui sentimen dari setiap komentar. Analisis sentimen merupakan suatu cabang ilmu dari text mining yang digunakan untuk mengekstrak, memahami, dan mengolah data teks. Untuk mengetahui setiap sentimen pada komentar digunakan fitur Term Frequency-Inverse Document Frequency (TF-IDF) dan beberapa metode klasifikasi seperti Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Gradient Boosting, Multinomial Naive Bayes. Dataset didapatkan dari kaggle dengan judul “*Cyberbullying Bahasa Indonesia*” yang berisi 650 sample komentar. Dokumen komentar dibagi menjadi 80% data latih dan 20% data uji. Berdasarkan pengujian yang dilakukan didapatkan hasil terbaik pada metode Random Forest dengan akurasi sebesar 97%, precision sebesar 94,44%, 97% recall. dan f1-score sebesar 89,47%. Untuk mempermudah akses pengguna maka model tersebut akan diimplementasikan ke dalam sebuah website.

Kata kunci: media sosial, *cyberbullying*, analisis sentimen, klasifikasi.

ABSTRACT

The usage of the internet in Indonesia continues to increase every year, with 215 million inhabitants recorded using the internet in 2023. Alongside the growing internet usage, various social media platforms like Facebook, Twitter, and Instagram are also increasingly popular among the public. However, these social media platforms are not immune to the dangers of *cyberbullying*, which is often done by users, especially in the comments sections. The danger of *cyberbullying* is of course troubling many people because of the impact it has, so one effort to prevent *bullying* on social media is to do a sentiment analysis in the social media comment sections which aims to find out the sentiment of each comment. Sentiment analysis is a branch of text mining that involves extracting, understanding, and processing textual data. To identify the sentiment in comments, the Term Frequency-Inverse Document Frequency (TF-IDF) feature and several classification methods, including Logistic Regression, Decision Tree, Random Forest, K-Nearest Neighbors, Gradient Boosting, and Multinomial Naive Bayes, are utilized. The dataset used for the analysis is obtained from Kaggle, titled "*Cyberbullying Bahasa Indonesia*," consisting of 650 sample comments. The comment documents are divided into 80% training data and 20% test data. Based on the conducted testing, the best results were achieved with the Random Forest method, showing an accuracy of 97%, precision of 94.44%, recall of 97%, and an f1-score of 89.47%. To enhance user accessibility, the model will be implemented into a website.

Keyword: social media, *cyberbullying*, sentiment analysis, classification