

**KLASIFIKASI CUITAN PADA MEDIA SOSIAL TWITTER
DENGAN LABEL BERDASARKAN PASAL KUHP DAN UU
ITE MENGGUNAKAN METODE C4.5**

SKRIPSI

Diajukan Untuk Memenuhi Sebagai Persyaratan Dalam Menempuh Gelar
Sarjana Komputer Program Studi Informatika



Oleh :

BENAYA ARTA HENDRATRI

NPM. 18081010037

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”

JAWA TIMUR

2022

LEMBAR PENGESAHAN

Judul : KLASIFIKASI CUITAN PADA MEDIA SOSIAL TWITTER
DENGAN LABEL BERDASARKAN PASAL KUHP DAN UU ITE
MENGGUNAKAN METODE C4.5

Oleh : BENAYA ARTA HENDRATRI

NPM : 18081010037

Telah Diseminarkan Dalam Ujian Skripsi Pada
Jumat, 24 Juni 2022
Mengetahui

Dosen Pembimbing

1.

Intan Yuniar Purbasari, S.Kom, M.Sc

NPT : 3 8006 04 0198 1

2.

Farwaz Ali Akbar, S.Kom, M.Kom

NIP : 19920317 201803 002

Dosen Pengaji

1.

Yisti Vita Vira, S.St, M.Kom

NIPPK : 19860425 2021212 001

2.

Eva Yulia P., S.Kom, M.Kom

NIPPK : 19890705 2021211 002

Menyetujui

Koordinator Program Studi

Informatika



Budi Nugroho, S.Kom., M.Kom

NIPPK : 19800907 202121 1 005

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Informatika UPN “Veteran” Jawa Timur, yang bertandatangan dibawah ini :

Nama : BENAYA ARTA HENDRATRI

NPM : 18081010037

Menyatakan bahwa judul skripsi / tugas akhir yang saya ajukan dan kerjakan, yang berjudul

**“ KLASIFIKASI CUITAN PADA MEDIA SOSIAL TWITTER DENGAN
LABEL BERDASARKAN PASAL KUHP DAN UU ITE
MENGGUNAKAN METODE C4.5”**

Bukan merupakan plagiat dari skripsi / tugas akhir / penelitian orang lain dan juga bukan merupakan produk atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa skripsi / tugas akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam daftar pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka saya siap menerima segala konsekuensinya.

Surabaya, 24 Juni 2022

Hormat Saya,



Benaya Arta Hendratri

NPM. 18081010037

KLASIFIKASI CUITAN PADA MEDIA SOSIAL TWITTER DENGAN LABEL BERDASARKAN PASAL KUHP DAN UU ITE MENGGUNAKAN METODE C4.5

Nama Mahasiswa : Benaya Arta Hendratri

NPM : 18081010037

Program Studi : Informatika

Dosen Pembimbing : 1. Intan Yuniar Purbasari, S.Kom, M.Sc

2. Fawwaz Ali Akbar, S.Kom, M.Kom

ABSTRAK

Media sosial Twitter menjadi salah satu alat dalam menyampaikan pendapat baik berupa saran ataupun kritik. Namun masyarakat belum dapat membedakan perbuatan yang melanggar peraturan-peraturan tersebut sehingga diperlukan filter yang dapat menjadi pengingat sebelum pengguna mengirim atau melakukan sesuatu di media sosial. Filter yang dibuat telah disesuaikan dengan UU ITE dan KUHP.

Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik.

Berdasarkan pengujian dari beberapa skenario uji coba nilai akurasi tertinggi yang dihasilkan sebesar 71,4% dengan skenario penggunaan data latih sebesar 80% dan data uji 10% serta fitur yang diambil sebesar 10%. Skenario ini memiliki *recall* sebesar 64,2%, *precision* 50% dan F1-score 56,2%. Dari perolehan tersebut dapat diambil kesimpulan jika metode C4.5 dapat digunakan sebagai metode klasifikasi pada penelitian berjudul Klasifikasi Cuitan Pada Media Sosial Twitter Dengan Label Berdasarkan Pasal KUHP Dan UU ITE Menggunakan Metode C4.5

Kata Kunci: *Twitter, UU ITE, KUHP, Algoritma C4.5*

UCAPAN TERIMAKASIH

Puji syukur kehadirat Allah SWT berkat ridho, taufik dan hidayah-Nya penulis dapat menyelesaikan laporan skripsi ini dengan tuntas dan maksimal. Dengan diselesaiannya laporan skripsi ini penulis ingin mengucapkan terimakasih kepada pihak-pihak yang telah berpartisipasi dalam penyusunan laporan skripsi ini, karena tanpa bantuan, saran, kritik serta dukungan mereka mungkin penulis tidak dapat menyelesaikan laporan ini.

Pada kesempatan ini penulis ingin mengucap syukur dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Ahmad Fauzi, M.MT. selaku Rektor UPN “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Ni Ketut Sari, MT. selaku Dekan Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur.
3. Bapak Budi Nugroho, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur
4. Ibu Intan Yuniar Purbasari, S.Kom, M.Sc selaku Dosen Pembimbing 1 yang telah memberikan kesempatan, ilmu serta pengetahuan kepada penulis dalam menyelesaikan laporan skripsi dengan maksimal.
5. Bapak Fawwaz Ali Akbar, S.Kom, M.Kom selaku Dosen Pembimbing 2 yang telah meluangkan waktu, pikiran dalam membimbing penulis sehingga laporan skripsi ini dapat terselesaikan dengan tuntas.
6. Bapak dan Ibu Dosen Program Studi Informatika UPN “Veteran” Jawa Timur yang telah memberikan arahan, ilmu, dan pembelajaran selama perkuliahan.
7. Kedua orang tua yaitu, Siti Maymunah dan Hendry Amin yang selalu memberikan dukungan do'a dan motivasi kepada penulis untuk selalu tetap semangat dan berusaha dalam meraih hasil maksimal dalam setiap langkah yang dilalui penulis.
8. Teman-teman Bumi Manusia yang selalu memberi motivasi dan bantuan ketika penulis sedang kesusahan dalam mengerjakan penelitian dan tugas-tugas pada masa perkuliahan.

9. Teman-teman Wiji Family yang selalu memberi dukungan dan bantuan ketika penulis sedang kesusahan dalam mengerjakan penelitian dan tugas-tugas pada masa perkuliahan.

Terima kasih untuk semua pihak yang telah membantu, membimbing dan memberikan motivasi pengerajan skripsi ini. Semoga Tuhan Yang Maha Esa memberikan kebaikan serta berkah untuk semua pihak yang berpartisipasi pada penyelesaian tugas akhir penulis. Penulis menyadari jika masih banyak kekurangan dalam penulisan laporan skripsi ini, oleh karena itu, saran dan kritik dari pembaca dapat membantu dalam menyempurnakan laporan skripsi ini. Penulis berharap semoga laporan skripsi ini dapat bermanfaat bagi pembaca atau penulis tersendiri. Akhir kata yang dapat saya sampaikan hanya terima kasih, penulis berharap semoga Tuhan Yang Maha Esa membalas segala kebaikan yang telah diberikan untuk semuanya. Aamiin.

KATAPENGANTAR

Syukur Alhamdulillah kami panjatkan kehadirat Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Cuitan Pada Media Sosial Twitter dengan Label Berdasarkan Pasal KUHP dan UU ITE Menggunakan Metode C4.5“ ini dengan baik dan lancar, Tujuan skripsi ini dibuat sebagai salah satu persyaratan untuk mendapatkan gelar Sarjana Komputer di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis berharap dengan penyusunan skripsi ini dapat memberikan manfaat bagi temanteman maupun pembaca pada umumnya. Penulis menyadari dalam penulisan skripsi ini masih jauh dari sempurna, dan banyak kekurangan baik dalam pembahasan materi maupun dalam penyusunan skripsi ini, sehingga penulis sangat mengharapkan kritik dan saran yang bersifat membangun yang nantinya dapat dijadikan pelajaran oleh penulis maupun pembaca dikemudian hari.

Tidak lupa penulis mengucapkan banyak terimakasih atas bantuan dari berbagai pihak yang telah berkontribusi dengan memberikan sumbangan baik materi maupun pikirannya Semoga laporan ini dapat bermanfaat bagi semua baik bagi penulis dan para pembaca pada umumnya.

Surabaya, 27 Juni 2022

Benaya Arta Hendratri

DAFTAR ISI

LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN ANTI PLAGIAT.....	iii
ABSTRAK	iv
UCAPAN TERIMAKASIH.....	v
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	3
1.5. Manfaat Penelitian.....	3
1.5.1. Bagi Penulis	3
1.5.2. Bagi Pengguna	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terdahulu.....	5
2.2. Twitter	7
2.3. Cuitan	7
2.4. UU ITE	7
2.5. KUHP	8
2.6. Klasifikasi.....	9
2.7. Metode C4.5	9
2.10.1 Perhitungan <i>Gain</i>	10
2.10.2 Pembentukan Pohon Keputusan.....	10
2.8. <i>Preprocessing</i>	11
2.8.1. <i>Tokenizing</i>	12

2.8.2. <i>CASEFOLDING</i>	12
2.8.3. Normalisasi	13
2.8.4. <i>Filtering</i>	13
2.8.5. <i>Stemming</i>	14
2.9. TF-IDF.....	15
2.10. <i>Confusion Matrix</i>	15
2.11. Phyton.....	16
2.12. Jupyter Notebook	16
2.13. Sastrawi	17
2.14. Numpy	17
2.15. Pandas.....	17
2.16. Kaggle	17
2.17. Regex.....	17
2.18. CSV	18
BAB III METODOLOGI PENELITIAN	19
3.1. Analisis Permasalahan.....	19
3.2. Analisis Data	20
3.3. Analisis Proses	22
3.3.1. <i>Preprocessing</i> Data	22
3.3.2. TF-IDF	25
3.3.3. Algoritma C4.5.....	26
3.4. Skenario Uji Coba	30
BAB IV HASIL DAN PEMBAHASAN	31
4.1. Pengumpulan Data	31
4.2. <i>Preprocessing</i>	31
4.2.1. <i>Tokenizing</i>	32
4.2.2. <i>CASEFOLDING</i>	33
4.2.3. Normalisasi	34
4.2.4. <i>Filtering</i>	35

4.2.5. <i>Stemming</i>	37
4.3. TF-IDF.....	38
4.4. Pembagian Data.....	42
4.5. Algoritma C4.5	42
4.6. Hasil Dan Pengujian	51
BAB V PENUTUP.....	69
5.1. Kesimpulan.....	69
5.2. Saran	69
DAFTAR PUSTAKA	71
LAMPIRAN	74
BIODATA PENULIS	96

DAFTAR GAMBAR

Gambar 3.1 Tahapan penelitian	19
Gambar 3.2 Hasil analisis proses	22
Gambar 3.3 Alur TF-IDF	26
Gambar 3.4 Alur algoritma C4.5.....	27
Gambar 3.5 Pohon keputusan skenario data 80%/20% dengan fitur 10%	29
Gambar 4.1 Implementasi data <i>frame</i> pada data mentah	31
Gambar 4.2 Implementasi data <i>frame</i> pada data mentah	32
Gambar 4.3 Hasil <i>Tokenizing</i> pada program.....	33
Gambar 4.4 Hasil <i>CASEFOLDING</i> pada program.....	34
Gambar 4.5 Kamus Normalisasi	34
Gambar 4.6 Hasil Normalisasi pada program.....	35
Gambar 4.7 <i>Stopword</i> tambahan	35
Gambar 4.8 Hasil <i>Filtering</i> pada program.....	36
Gambar 4.9 Hasil <i>Stemming</i> pada program.....	37
Gambar 4.10 Hasil perhitungan TF.....	39
Gambar 4.11 Jenis kata	40
Gambar 4.12 Hasil perhitungan IDF.....	41
Gambar 4.13 Hasil perhitungan TF-IDF.....	41
Gambar 4.14 Pohon keputusan skenario 1	45
Gambar 4.15 Pohon keputusan skenario 2	45
Gambar 4.16 Klasifikasi dengan model skenario 1	51
Gambar 4.17 Hasil klasifikasi dengan model skenario 1	52
Gambar 4.18 Klasifikasi dengan model skenario 2	52
Gambar 4.19 Hasil klasifikasi dengan model skenario 2	52
Gambar 4.20 Klasifikasi dengan model skenario 3	53
Gambar 4.21 Hasil klasifikasi dengan model skenario 3	53
Gambar 4.22 Klasifikasi dengan model skenario 4	54
Gambar 4.23 Klasifikasi dengan model skenario 4	54
Gambar 4.24 Klasifikasi dengan model skenario 5	54
Gambar 4.25 Hasil klasifikasi dengan model skenario 5	55
Gambar 4.26 Klasifikasi dengan model skenario 6	55

Gambar 4.27 Hasil klasifikasi dengan model skenario 6	56
Gambar 4.28 Klasifikasi dengan model skenario 7	56
Gambar 4.29 Hasil klasifikasi dengan model skenario 7	56
Gambar 4.30 Klasifikasi dengan model skenario 8	57
Gambar 4.31 Hasil klasifikasi dengan model skenario 8.....	57
Gambar 4.32 Klasifikasi dengan model skenario 9	58
Gambar 4.33 Hasil klasifikasi dengan model skenario 9	58
Gambar 4.34 Klasifikasi dengan model skenario 10	59
Gambar 4.35 Hasil klasifikasi dengan model skenario 10.....	59
Gambar 4.36 Klasifikasi dengan model skenario 11	60
Gambar 4.37 Hasil klasifikasi dengan model skenario 11	60
Gambar 4.38 Klasifikasi dengan model skenario 12	61
Gambar 4.39 Hasil klasifikasi dengan model skenario 12.....	61
Gambar 4.40 Model C4.5 dengan perolehan akurasi terburuk	62
Gambar 4.41 Grafik akurasi setiap persentase fitur	68
Gambar 4.42 Rata – rata persentase pada setiap fitur	68

DAFTAR TABEL

Tabel 2.1 Proses <i>Tokenizing</i>	12
Tabel 2.3 Proses <i>CASEFOLDING</i>	13
Tabel 2.2 Proses Normalisasi	13
Tabel 2.4 Proses <i>Filtering</i>	14
Tabel 2.5 Proses <i>Stemming</i>	14
Tabel 3.1 Atribut data	20
Tabel 3.2 Inisial kelas pada dataset.....	23
Tabel 3.3 <i>Tokenizing</i>	23
Tabel 3.4 <i>CASEFOLDING</i>	23
Tabel 3.5 Normalisasi	24
Tabel 3.6 <i>Filtering</i>	24
Tabel 3.7 <i>Stemming</i>	25
Tabel 3.8 Skenario uji coba.....	30
Tabel 4.1 Akurasi setiap skenario	62
Tabel 4.2 Penentuan <i>Confussion Matrix</i> skenario 2.....	63
Tabel 4.3 <i>Confussion Matrix</i> skenario 2	63
Tabel 4.4 Perhitungan <i>precision, recall dan F1-score</i> skenario 2	63
Tabel 4.5 Penentuan <i>Confussion Matrix</i> skenario 3.....	64
Tabel 4.6 <i>Confussion Matrix</i> skenario 3	64
Tabel 4.7 Perhitungan <i>precision, recall dan F1-score</i> skenario 3	65
Tabel 4.8 Penentuan <i>Confussion Matrix</i> skenario 6.....	65
Tabel 4.9 <i>Confussion Matrix</i> skenario 6	66
Tabel 4.10 Perhitungan <i>precision, recall dan F1-score</i> skenario 6.....	66