

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
CUTAN UJARAN KEBENCIAN PADA TWITTER**

SKRIPSI



disusun oleh

Dikih Arif Wibowo

18.21.1158

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
CUTAN UJARAN KEBENCIAN PADA TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Dikih Arif Wibowo

18.21.1158

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
CUITAN UJARAN KEBENCIAN PADA TWITTER**

yang dipersiapkan dan disusun oleh

**Dikih Arif Wibowo
18.21.1158**

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 29 Agustus 2019

Dosen Pembimbing,



**Windha Mega Pradnya D, M.Kom.
NIK. 190302185**

PENGESAHAN
SKRIPSI
IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI
CUITAN UJARAN KEBENCIAN PADA TWITTER

yang dipersiapkan dan disusun oleh

Dikih Arif Wibowo

18.21.1158

telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Agustus 2019

Susunan Dewan Penguji

Nama Penguji

Kusnawi, S.Kom, M.Eng.
NIK. 190302112

Mei P Kurniawan, M.Kom.
NIK. 190302187

Ainul Yaqin, M.Kom.
NIK. 190302255

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 31 Agustus 2019

DEKAN FAKULTAS ILMU KOMPUTER



Krisnawati, S.Si., M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggung jawab saya pribadi.

Yogyakarta, 29 Agustus 2019



Arif Wibowo
Dikih Arif Wibowo
NIM. 18.21.1158

MOTTO

“Dadi opo wae kudu dadi wong seng : Berilmu Amaliyah Beramal Ilmiah

Bertaqwa Ilahiyah”

- KH. Aliy As’ad –

“Kanjeng Nabi iku Wong, sing ngerti wong, nguwongna wong.”

- KH. Mustofa Bisri (Gus Mus) –

“Gitu aja kok repot”

- KH. Abdurrahman Wahid (Gus Dur) -

“Kalau sampean serius mau memperjuangkan. Tunggu saya lulus pondok dan
Datanglah ke rumah dan matur ke bapa ibu”

- Anonymous -

PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung maupun tidak langsung dalam proses pembuatan skripsi.

1. Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayangnya sampai sejauh ini.
2. Kedua orang tua saya dan keluarga, yang selalu mendoakan, selalu menyemangati dan memberikan dukungan baik secara material maupun non material.
3. Ibu Windha Mega Pradnya, M.Kom yang telah membimbing saya dari awal sampai akhir pembuatan skripsi.
4. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
5. K.H. M. Rajif Dienal Maula, Lc., M.Ag. dan segenap keluarga ndalem Pondok Pesantren Nailul Ula Center yang saya hormati dan saya tunggu selalu petuah-petuhannya.
6. Teman-teman kelas D3 TI 01 2015 dan S1 IFT 2018 yang selalu menemani perkuliahan, mendukung dan memberikan semangat sampai saat ini. Semoga kita selalu bahagia dan menjadi pribadi yang lebih baik lagi.
7. Teman-teman dari Pondok Pesantren Nailul Ula Center Plosokuning sebagai teman ngaji dan media melepas lelah kala selesai perkuliahan.
8. Teman-teman ORMA Amikom Computer Club (AMCC) sebagai tempat sharing dan teman makan di angkringan.

KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Tuhan Yang Maha Esa yang telah memberikan rahmat, hidayah dan kekuatan sehingga saya dapat menyelesaikan skripsi yang berjudul Implementasi Algoritma *Naive Bayes* untuk Klasifikasi Cuitan Ujaran Kebencian pada Twitter.

Skripsi ini saya buat guna menyelesaikan studi jenjang Strata Satu (S1) pada program studi Informatika fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta. Selain itu juga merupakan suatu bukti bahwa mahasiswa telah menyelesaikan kuliah jenjang program strata satu dan untuk memperoleh gelar Sarjana Komputer. Dengan selesainya skripsi ini, maka pada kesempatan ini saya mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Ibu Windha Mega Pradnya, M.Kom. selaku dosen pembimbing yang selalu bijaksana memberikan bimbingan, nasehat serta waktunya selama penulisan skripsi ini.
4. Dosen Penguji dan segenap Dosen serta Karyawan Universitas AMIKOM Yogyakarta yang telah berbagi ilmu dan pengalamannya.
5. Kedua orang tua saya yang telah mendoakan, mendukung dan memberikan semangat kepada saya.
6. K.H. M. Rajif Dienal Maula Lc., M.Ag. dan segenap keluarga ndalam Pondok Pesantren Nailul Ula Center Plosokuning yang dengan penuh keikhlasan dan tanpa lelah memberikan ilmu agama dan membimbing saya agar menjadi manusia yang berguna bagi masyarakat.
7. Orang-orang baik yang senantiasa mendoakan, memberi semangat dan bantuan kepada saya.
8. Teman-teman D3 TI 01 angkatan 2015 dan S1 IFT 01 yang telah menemani selama proses perkuliahan.

9. Teman-teman Pondok Pesantren Nailul Ula Center Plosokuning yang telah menjadi teman ngaji dan memberikan semangat kepada saya.
10. Teman-teman Amikom Computer Club yang telah menjadi tempat belajar tentang organisasi dan teknologi.
11. Semua pihak yang telah membantu baik secara langsung maupun tidak langsung.

Semoga Allah subhanahu wa ta'ala memberikan balasan yang lebih kepada semua yang telah ikut membantu saya hingga menyelesaikan skripsi ini. Demi perbaikan selanjutnya, saran dan kritik yang membangun diterima dengan senang hati dan rasa terima kasih. Semoga skripsi ini dapat bermanfaat bagi saya dan kita semua.

Yogyakarta, 20 Agustus 2019



Dikih Arif Wibowo

DAFTAR ISI

JUDUL	I
PERSETUJUAN	II
PENGESAHAN	III
PERNYATAAN.....	IV
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI.....	IX
DAFTAR TABEL.....	X1
DAFTAR GAMBAR	XII
INTISARI.....	XIV
<i>ABSTRACT</i>	XV
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH	2
1.3 BATASAN MASALAH	3
1.4 MAKSUD DAN TUJUAN PENELITIAN	3
1.5 METODE PENELITIAN	4
1.6 SISTEMATIKA PENULISAN	5
BAB II LANDASAN TEORI.....	7
2.1 KAJIAN PUSTAKA	7
2.2 DASAR TEORI.....	11
2.2.1 Data Mining	11
2.2.2 Metode <i>Naive Bayes</i>	13
2.2.3 Ujaran Kebencian.....	15
2.2.4 <i>K-Fold Cross Validation</i>	16

2.2.5	<i>Data Flow Diagram(DFD)</i>	16
2.2.7	<i>Stemming</i>	18
2.2.8	Klasifikasi	18
2.2.9	<i>Confusion Matrix</i>	18
BAB III ANALISIS DAN PERANCANGAN		22
3.1	PROFIL PERUSAHAAN TWITTER	22
3.2	ANALISIS MASALAH.....	23
3.3	HASIL ANALISIS	24
3.4	DESKRIPSI SISTEM.....	24
3.5	PERHITUNGAN MANUAL NAIVE BAYES.....	26
3.5.1	Proses Training	26
3.5.2	Proses Klasifikasi.....	42
3.6	ANALISIS KEBUTUHAN PERANGKAT LUNAK.....	44
3.6.1	Definisi Fitur.....	44
3.7	DFD (DATA FLOW DIAGRAM).....	45
3.8	RANCANGAN ANTARMUKA PENGGUNA (<i>USER INTERFACE</i>).....	51
BAB IV IMPLEMENTASI DAN PEMBAHASAN		57
4.1	DATASET TRAINING	57
4.2	IMPLEMENTASI ANTARMUKA PENGGUNA.....	58
4.3	PEMBAHASAN <i>SOURCE CODE</i>	63
4.4	HASIL DAN ANALISIS PENGUJIAN MODEL	66
4.4.1	Pengujian Menggunakan 10-Fold Cross Validation dan Confusion Matrix	66
4.4.2	Hasil Pengujian	68
BAB V PENUTUP		92
5.1	KESIMPULAN	92
5.2	SARAN	92
DAFTAR PUSTAKA		93

DAFTAR TABEL

Tabel 2.1 Perbedaan Penelitian	9
Tabel 2.2 <i>Data Flow Diagram(DFD)</i>	166
Tabel 2.3 <i>Confusion Matrix</i>	19
Tabel 3.1 <i>Data Training</i>	26
Tabel 3.2 Jumlah Dokumen	35
Tabel 3.3 Frekuensi Token.....	36
Tabel 3.4 Daftar Fitur Aplikasi.....	44
Tabel 4.1 Hasil <i>Confusion Matrix Fold 1</i>	70
Tabel 4.2 Hasil <i>Confusion Matrix Fold 2</i>	72
Tabel 4.3 Hasil <i>Confusion Matrix Fold 3</i>	74
Tabel 4.4 Hasil <i>Confusion Matrix Fold 4</i>	76
Tabel 4.5 Hasil <i>Confusion Matrix Fold 5</i>	78
Tabel 4.6 Hasil <i>Confusion Matrix Fold 6</i>	80
Tabel 4.7 Hasil <i>Confusion Matrix Fold 7</i>	82
Tabel 4.8 Hasil <i>Confusion Matrix Fold 8</i>	84
Tabel 4.9 Hasil <i>Confusion Matrix Fold 9</i>	86
Tabel 4.10 Hasil <i>Confusion Matrix Fold 10</i>	88
Tabel 4.11 Hasil Pengujian Model Menggunakan <i>10 Fold Cross Validation</i> dan <i>Confusion Matrix</i>	88
Tabel 4.12 Hasil Pengujian Menggunakan <i>Confusion Matrix</i> Secara Keseluruhan	91

DAFTAR GAMBAR

Gambar 3.1 Diagram Alir Klasifikasi	25
Gambar 3.2 Gambar Diagram Konteks.....	45
Gambar 3.3 DFD(<i>Data Flow Diagram</i>) Level 1	46
Gambar 3.4 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Input Dataset.....	47
Gambar 3.5 DFD (<i>Data Flow Diagram</i>) Level 2 Proses <i>Training</i>	47
Gambar 3.6 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Prediksi.....	48
Gambar 3.7 DFD (<i>Data Flow Diagram</i>) Level 2 Proses Validasi.....	48
Gambar 3.8 DFD (<i>Data Flow Diagram</i>) Level 3 Proses <i>Training</i>	49
Gambar 3.9 DFD(<i>Data Flow Diagram</i>) Level 3 Proses Prediksi.....	50
Gambar 3.10 DFD (<i>Data Flow Diagram</i>) Level 3 Proses Validasi.....	51
Gambar 3.11 Main Menu	52
Gambar 3.12 Halaman Prediksi	53
Gambar 3.13 Halaman Latih Model.....	54
Gambar 3.14 Halaman Validasi	55
Gambar 3.15 Halaman Hasil Validasi.....	56
Gambar 4.1 <i>Dataset Hatespeech</i>	58
Gambar 4.2 <i>Dataset Nothatespeech</i>	58
Gambar 4.3 Implementasi Main Menu	59
Gambar 4.4 Implementasi Halaman Prediksi	60
Gambar 4.5 Implementasi Halaman Latih Model.....	60
Gambar 4.6 Implementasi Halaman Validasi	61
Gambar 4.7 Implementasi Halaman Hasil Validasi.....	62
Gambar 4.8 <i>Source Code</i> Pengambilan Data Komentar Dari Twitter.....	63
Gambar 4.9 <i>Source Code</i> Latih Model	64
Gambar 4.10 <i>Source Code Text Preprocessing</i>	65
Gambar 4.11 <i>Source Code</i> Proses Klasifikasi	66
Gambar 4.12 Skenario Pengujian <i>10-Fold Cross Validation</i>	67
Gambar 4.13 Hasil Pengujian <i>Fold 1</i>	68
Gambar 4.14 <i>Pie Chart</i> Hasil Pengujian <i>Fold 1</i>	69
Gambar 4.15 Hasil Pengujian <i>Fold 2</i>	70

Gambar 4.16 <i>Pie Chart</i> Hasil Pengujian <i>Fold 2</i>	71
Gambar 4.17 Hasil Pengujian <i>Fold 3</i>	72
Gambar 4.18 <i>Pie Chart</i> Hasil Pengujian <i>Fold 3</i>	73
Gambar 4.19 Hasil Pengujian <i>Fol 4</i>	74
Gambar 4.20 <i>Pie Chart</i> Hasil Pengujian <i>Fold 4</i>	75
Gambar 4.21 Hasil Pengujian <i>Fold 5</i>	76
Gambar 4.22 <i>Pie Chart</i> Hasil Pengujian <i>Fold 5</i>	77
Gambar 4.23 Hasil Pengujian <i>Fold 6</i>	78
Gambar 4.24 <i>Pie Chart</i> Hasil Pengujian <i>Fold 6</i>	79
Gambar 4.25 Hasil Pengujian <i>Fold 7</i>	80
Gambar 4.26 <i>Pie Chart</i> Hasil Pengujian <i>Fold 7</i>	81
Gambar 4.27 Hasil Pengujian <i>Fold 8</i>	82
Gambar 4.28 <i>Pie Chart</i> Hasil Pengujian <i>Fold 8</i>	83
Gambar 4.29 Hasil Pengujian <i>Fold 9</i>	84
Gambar 4.30 <i>Pie Chart</i> Hasil Pengujian <i>Fold 9</i>	85
Gambar 4.31 Hasil Pengujian <i>Fold 10</i>	86
Gambar 4.22 <i>Pie Chart</i> Hasil Pengujian <i>Fold 10</i>	87
Gambar 4.23 Grafik Akurasi Model	90

INTISARI

Twitter merupakan media sosial yang populer di Indonesia. Di twitter pengguna bisa membuat cuitan dalam bentuk text, foto atau video. Para pengguna twitter dapat saling mengikuti satu sama lain, membuat cuitan, membalas cuitan dan membagikan kembali cuitan. Di twitter pengguna dalam cuitannya dapat membuat tagar, tagar yang sedang populer akan menjadi peringkat teratas di twitter.

Namun seringkali banyak cuitan yang mengandung ujaran kebencian dalam bentuk hinaan, berita bohong yang ditujukan kepada orang, ras, agama dan kelompok tertentu. Seringkali cuitan ujaran kebencian dibuat oleh akun palsu. Twitter telah membuat fitur yang memungkinkan pengguna untuk melaporkan akun yang terindikasi sebagai bot atau akun palsu, akan tetapi masih diperlukan model untuk mendeteksi cuitan tersebut mengandung ujaran kebencian atau tidak.

Algoritma *Naive Bayes* akan mencari probabilitas setiap kelas ketika cuitan yang akan di prediksi di inputkan. Sebelum cuitan di hitung nilai probabilitasnya pada masing-masing kelas, cuitan akan diproses melalui tahap *preprocessing* yaitu *casefolding*, *cleaning*, *tokenizing*, dan *stemming*. Setelah diketahui nilai probabilitas dari masing-masing kelas, selanjutnya akan dibandingkan nilai probabilitasnya. Jika nilai probabilitasnya yang tertinggi adalah cuitan yang dicek dengan hipotesa kelas cuitan ujaran kebencian. Maka cuitan itu akan dilabeli sebagai sebagai cuitan ujaran kebencian. Jika nilai probabilitasnya yang tertinggi adalah cuitan yang dicek dengan hipotesa kelas cuitan tidak ujaran kebencian, maka cuitan tersebut akan dilabeli sebagai cuitan tidak mengandung ujaran kebencian.

Kata kunci: *Naive Bayes, Bayesian, Text Classification, Preprocessing, Twitter.*

ABSTRACT

Twitter is a popular social media in Indonesia. On Twitter, users can make text in the form of text, photos or videos. Twitter users can follow each other, make tweets, reply and share again. On Twitter users in their tweet can make hashtags, popular hashtags will be ranked top on Twitter.

However, there are often many stories that contain utterances of hatred in the form of insults, false news directed at people, races, religions and certain groups. Often cuitan utterances of hatred are made by fake accounts. Twitter has created a feature that allows users to report accounts indicated as bots or fake accounts, but a model is still needed to detect whether the account contains hate speech or not.

The Naive Bayes algorithm will look for the probability of each class when the input that is predicted is inputted. Before cuitan calculated the probability value in each class, cuitan will be processed through the preprocessing stage, namely casefolding, cleaning, tokensizing, and stemming. After knowing the probability value of each class, it will then compare the probability value. If the highest probability value is a trick which is checked by the hypothesis that the class is a speech of hatred. Then the cuitan will be labeled as a speech of hate. If the highest probability value is a score checked by the hypothesis that the class is not a statement of hate, then the statement will be labeled as cuitan not containing utterances of hatred.

Keyword: *Naive Bayes, Bayesian, Text Classification, Preprocessing, Twitter.*