

**KLASIFIKASI UJARAN KEBENCIAN PADA CUITAN  
DALAM BAHASA INDONESIA**

**TUGAS AKHIR**

**Diajukan untuk Memenuhi Salah Satu Persyaratan Mencapai Derajat  
Sarjana Teknik Informatika**



Dibuat Oleh:

**Kevin Antariksa**

**150708214**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ATMA JAYA YOGAKARTA  
2019**

**LEMBAR PENGESAHAN**  
**KLASIFIKASI UJARAN KEBENCIAN PADA CUITAN DALAM BAHASA**  
**INDONESIA**

Yogyakarta, 25 Juli 2019

Kevin Antariksa

150708214

Menyetujui,

Pembimbing I

Y. Sigit Purnomo W.P., S.T., M.Kom.

NPP : 04.01.710

Pembimbing II

Dra. Ernawati, M.T.

NPP : 09.92.407

Penguji I

Y. Sigit Purnomo W.P., S.T., M.Kom.

NPP : 04.01.710

Penguji II

Dr. Pranowo, S.T., M.T.


NPP : 09.96.596

Penguji III

Eddy Julianito, S.T., M.T.

NPP : 07.97.624

Mengetahui,

 Dekan Fakultas Teknologi Industri



Dr. A. Teguh Siswanto, M.Sc.

NPP : 09.93.464

## **Pernyataan Orisinalitas & Publikasi Ilmiah**

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Kevin Antariksa  
NPM : 150708214  
Program Studi : Teknik Informatika  
Fakultas : Teknologi Industri  
Judul Penelitian : Klasifikasi Ujaran Kebencian pada Cuitan dalam Bahasa Indonesia

Menyatakan dengan ini:

1. Tugas Akhir ini adalah benar tidak merupakan salinan sebagian atau keseluruhan dari karya penelitian lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta atas penelitian ini, berupa Hak untuk menyimpan, mengelola, mendistribusikan, dan menampilkan hasil penelitian selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum atas pelanggaran Hak Cipta dalam pembuatan Tugas Akhir ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 09 Juli 2019

Yang menyatakan,



Kevin Antariksa

15 07 08214

## HALAMAN PERSEMBAHAN

*“To everything there is a season, and a time to every purpose under the heaven. A time to weep, and a time to laugh. A time to get, and a time to lose. A time to keep, and a time to cast away.*

***He hath made everything beautiful in His time.”***

*Ecclesiastes 3*

*A tribute to my beloved family.*

## KATA PENGANTAR

Puji Syukur dan terima kasih penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir ini. Penyusunan tugas akhir ini tidak lepas dari bantuan berbagai pihak yang senantiasa mendukung penulis. Oleh sebab itu penulis memberikan ucapan terima kasih kepada :

1. Tuhan Yesus Kristus, yang selalu menyertai dan melimpahkan rahmat dan karunia-Nya kepada penulis.
2. Papa, Mama, Rafael, dan seluruh keluarga, yang selalu mendukung penulis.
3. Bapak Y. Sigit Purnomo selaku dosen pembimbing I yang selama ini telah membimbing penulis selama penulisan dan pengerjaan tugas akhir ini.
4. Ibu Ernawati selaku dosen pembimbing II yang telah meneliti dan membantu penulisan tugas akhir ini.
5. Bapak Martinus Maslim selaku kepala prodi Teknik Informatika dan seluruh dosen prodi Teknik Informatika UAJY yang telah mendukung proses perkuliahan.
6. Teman-teman PG (Wahyu, Dika, Bobby, Daniel, Komang, Ian, Verry, Travis) yang telah menghibur, memberikan semangat kepada penulis untuk melaksanakan dan menyelesaikan tugas akhir ini.
7. Teman-teman kos TB V 5A (Ronald, Samuel, Vicky, Aswin, Nanda) yang selalu mendukung penulis.
8. Sr. Teresa Tri Astuti yang telah membimbing, memotivasi, menjadi tempat *sambat* dan memberikan solusi kepada saya selama 4 tahun berkuliah.
9. Teman-teman komunitas PSSB – BIDIK MISI beserta seluruh pengurus harian PSSB – BIDIK MISI yang telah menemani saya selama berproses selama 4 tahun di Atma Jaya ini.
10. Teman-teman magang ACC (Elen, Budi, Valen, Desya, Rani, Nugie, Wira, Rexy, Teddy) yang memberikan semangat dan telah membantu selama magang.

11. Teman-teman HIMAFORKA, teman-teman KSP, teman-teman Asisten ALPRO, Asisten KECEBUT, Asisten DASPRO, dan Asisten PAW yang menemani berproses selama masa perkuliahan.
  12. Seluruh teman-teman dan saudara yang telah memberikan bantuan serta dukungan pada penulis yang tidak dapat penulis sebutkan.
- Penulis berharap laporan ini akan memberikan manfaat kepada pembaca.

Yogyakarta, 16 Juni 2019

Penulis,  
Kevin Antariksa

## DAFTAR ISI

LEMBAR PENGESAHAN .....	i
Pernyataan Orisinalitas & Publikasi Ilmiah.....	ii
HALAMAN PERSEMBAHAN .....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
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    Sistematika Penelitian .....	3
BAB II TINJAUAN PUSTAKA .....	5
BAB III LANDASAN TEORI.....	10
3.    Landasan Teori.....	10
3.1    Ujaran kebencian.....	10
3.2    Pembelajaran Mesin.....	12
3.3    Algoritma SVM .....	13
3.4    Algoritma Naïve Bayes.....	14
3.5    Algoritma Logistic Regression.....	15
3.6    TF-IDF .....	15
3.7    N-grams.....	16
3.8    Word2vec.....	16
BAB IV METODE PENELITIAN .....	18
4.1    Metode Studi Pustaka .....	19
4.2    Metode Pembangunan Model Pembelajaran Mesin .....	19
4.2.1    Pembuatan dataset.....	19
4.2.2    Pembersihan dataset.....	20
4.2.3    Feature Extraction.....	20
4.2.4    Penerapan Algoritma.....	21
4.2.5    Penyesuaian Model ( <i>tuning</i> ).....	21
4.2.6    Evaluasi Model.....	22

<b>BAB V HASIL PENELITIAN DAN PEMBAHASAN</b>	23
<b>5.1    Pembuatan Dataset</b>	23
5.1.1    Pengumpulan Dataset	23
5.1.2    Pembersihan Dataset	25
5.1.3    Pemberian label	27
<b>5.2    Pembuatan Model</b>	28
5.2.1    Memuat Dataset	28
5.2.2    Feature Extraction	30
5.2.2.1    TF-IDF	30
5.2.2.2    N-gram	31
5.2.2.3    Word2Vec	31
5.2.3    Algoritma Klasifikasi	34
5.2.3.1    Penerapan Algoritma	34
<b>5.3    Evaluasi Model</b>	37
5.3.1    Naïve Bayes dengan TF-IDF	38
5.3.2    Naïve Bayes dengan N-gram	39
5.3.3    Naïve Bayes dengan Word2Vec	44
5.3.4    Logistic Regression dengan TF-IDF	45
5.3.5    Logistic Regression dengan N-gram	46
5.3.6    Logistic Regression dengan Word2Vec	48
5.3.7    SVM dengan TF-IDF	49
5.3.8    SVM dengan N-gram	50
5.3.9    SVM dengan Word2Vec	51
<b>5.4    Pembahasan</b>	52
<b>BAB VI KESIMPULAN DAN SARAN</b>	56
<b>6.1    Kesimpulan</b>	56
<b>6.2    Saran</b>	56
<b>Daftar Pustaka</b>	58

## DAFTAR GAMBAR

Gambar 4. 1 Diagram alir penelitian .....	18
Gambar 4. 2 Contoh consumer key .....	19
Gambar 5. 1 Fungsi pembuatan berkas untuk dataset .....	24
Gambar 5. 2 Fungsi pengambilan data-data dari Twitter.....	24
Gambar 5. 3 Menuliskan data yang telah diambil ke dalam berkas csv .....	25
Gambar 5. 4 Melakukan slice untuk memisahkan elemen-elemen dalam cuitan. 26	
Gambar 5. 5 Fungsi untuk memisahkan emoji dan emoticon dari dataset.....	26
Gambar 5. 6 Contoh tweet_text yang berisikan ujaran kebencian .....	27
Gambar 5. 7 Komposisi dataset yang diberi label.....	28
Gambar 5. 8 Dataset yang berhasil dimuat .....	29
Gambar 5. 9 Mengatur jumlah ‘lipatan’ pada K-Fold.....	29
Gambar 5. 10 Pemisahan menggunakan K-Fold.....	30
Gambar 5. 11 Penggunaan TF-IDF .....	30
Gambar 5. 12 Penggunaan N-gram .....	31
Gambar 5. 13 Pemisahan setiap kata untuk setiap baris .....	32
Gambar 5. 14 Penggabungan seluruh kata-kata .....	32
Gambar 5. 15 Melakukan feature extraction.....	32
Gambar 5. 16 Contoh hasil feature extraction.....	32
Gambar 5. 17 Fungsi untuk Word2Vec .....	33
Gambar 5. 18 Fungsi untuk melakukan prediksi.....	35
Gambar 5. 19 Fungsi untuk menampilkan hasil dari prediksi.....	35
Gambar 5. 20 Fungsi untuk plotting (penggambaran) confusion matrix .....	36
Gambar 5. 21 Penggunaan Naïve Bayes .....	36
Gambar 5. 22 Penggunaan SVM.....	37
Gambar 5. 23 Penggunaan Logistic Regression.....	37
Gambar 5. 24 Hasil prediksi menggunakan Multinomial Naïve Bayes dan TF-IDF .....	38
Gambar 5. 25 Hasil prediksi menggunakan Bernoulli Naïve Bayes dan TF-IDF.	38
Gambar 5. 26 Pengaturan pada N-gram .....	40

Gambar 5. 27 Hasil prediksi menggunakan Multinomial Naïve Bayes dengan word Level N-gram .....	41
Gambar 5. 28 Hasil prediksi menggunakan Multinomial Naïve Bayes dengan char Level N-gram .....	41
Gambar 5. 29 Hasil prediksi menggunakan Bernoulli Naïve Bayes dengan word Level N-gram .....	42
Gambar 5. 30 Hasil prediksi menggunakan Multinomial Naïve Bayes dengan char Level N-gram .....	43
Gambar 5. 31 Hasil prediksi menggunakan Multinomial Naïve Bayes dengan Word2Vec .....	44
Gambar 5. 32 Hasil prediksi menggunakan Bernoulli Naïve Bayes dengan Word2Vec .....	44
Gambar 5. 33 Hasil prediksi menggunakan Logistic Regression dengan TF-IDF	45
Gambar 5. 34 Hasil prediksi menggunakan Logistic Regression dengan Word Level N-gram .....	46
Gambar 5. 35 Hasil prediksi menggunakan Logistic Regression dengan char level N-gram .....	47
Gambar 5. 36 Hasil prediksi menggunakan Logistic Regression dengan Word2Vec .....	48
Gambar 5. 37 Hasil prediksi menggunakan SVM – TF-IDF .....	49
Gambar 5. 38 Hasil prediksi menggunakan SVM dan N-gram word level .....	50
Gambar 5. 39 Hasil prediksi menggunakan SVM dan N-gram char level.....	50
Gambar 5. 40 Hasil prediksi menggunakan SVM – Word2Vec .....	51

## **DAFTAR TABEL**

Tabel 2. 1 Tabel perbandingan dengan penelitian sebelumnya .....	8
Tabel 5. 1 Tabel perbandingan antara algoritma dan feature extraction .....	55

## **Intisari**

### **KLASIFIKASI UJARAN KEBENCIAN PADA CUITAN DALAM BAHASA INDONESIA**

Kevin Antariksa

150708214

Banyaknya ujaran kebencian yang ada di media sosial sudah membuat jengah. Ujaran kebencian tersebut makin marak dijumpai namun masih belum ada upaya preventif dari media sosial untuk menangkai ujaran kebencian. Deteksi ujaran kebencian yang sudah dibuat juga belum tersedia dalam Bahasa Indonesia.

Pada tugas akhir ini, akan dibuat model pembelajaran mesin yang dapat mengenali ujaran kebencian dengan Bahasa Indonesia. Dalam model tersebut akan membandingkan beberapa metode-metode pembelajaran mesin yang ada. Metode yang digunakan dalam pengujian adalah Naïve Bayes, SVM, dan Logistic Regression. Dalam pengujian, beberapa parameter akan diubah-ubah sehingga didapatkan nilai paling maksimal dalam deteksi ujaran kebencian.

Hasil yang diharapkan adalah sebuah model pembelajaran mesin. Model tersebut diharapkan dapat mengenali ujaran kebencian berbahasa Indonesia secara akurat. Akurasi yang diharapkan adalah sebesar >85%.

Kata kunci: Deteksi ujaran kebencian; Model pembelajaran mesin; media sosial; Bahasa Indonesia; cuitan;

Dosen Pembimbing I : Y. Sigit Purnomo W.P., S.T., M.Kom.

Dosen Pembimbing II : Dra. Ernawati, M.T

Jadwal Sidang Tugas Akhir : 24 Juli 2019