

**DETEKSI UJARAN KEBENCIAN BAHASA BANJAR MENGGUNAKAN
*ENSEMBLE METHOD***

TUGAS AKHIR

Oleh:

MUHAMMAD NUR ABDI

NIM. 1910817110008



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN

2023

**DETEKSI UJARAN KEBENCIAN BAHASA BANJAR MENGGUNAKAN
*ENSEMBLE METHOD***

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Sarjana Strata-1 Teknologi Informasi

Oleh:

MUHAMMAD NUR ABDI

NIM. 1910817110008



PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS TEKNIK

UNIVERSITAS LAMBUNG MANGKURAT

BANJARMASIN, JUNI 2023

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini,

Nama : Muhammad Nur Abdi

NIM : 1910817110008

Fakultas : Teknik

Program Studi : Teknologi Informasi

Judul Tugas Akhir : Deteksi Ujaran Kebencian Bahasa Banjar
Menggunakan *Ensemble Method*

Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom.

Pembimbing Pendamping : Muti'a Maulida, S.Kom., M.T.I

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi, dan sepanjang pengetahuan saya, juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, 31 Mei 2023

Muhammad Nur Abdi

NIM. 1910817110008

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN
SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
Deteksi Ujaran Kebencian Bahasa Banjar Menggunakan *Ensemble Method*
oleh
Muhammad Nur Abdi (1910817110008)

Telah dipertahankan di depan Tim Penguji pada 5 Juni 2023 dan dinyatakan

LULUS

Komite Penguji :

Ketua : Dr. Ir. Yuslena Sari, S.Kom., M.Kom.
NIP 198411202015042002

Anggota 1 : Eka Setya Wijaya, S.T., M.Kom.
NIP 198205082008011010

Anggota 2 : Nurul Fathanah Mustamin, S.Pd., M.T.
NIP 199110252019032018

Pembimbing : Andreyan Rizky Baskara, S.Kom., M.Kom.
Utama NIP 199307032019031011

Pembimbing : Muti'a Maulida, S.Kom., M.T.I.
Pendamping NIP 198810272019032013



Banjarbaru, 21 JUN 2023
diketahui dan disahkan oleh:

Wakil Dekan Bidang Akademik
Fakultas Teknik ULM,

Dr. Mahmud, S.T., M.T.
NIP 197401071998021001

Koordinator Program Studi
S-1 Teknologi Informasi,


Dr. Ir. Yuslena Sari, S.Kom., M.Kom.
NIP 198411202015042002

LEMBAR PERSETUJUAN

PERSETUJUAN TUGAS AKHIR

DETEKSI UJARAN KEBENCIAN BAHASA BANJAR MENGGUNAKAN
ENSEMBLE METHOD

OLEH
MUHAMMAD NUR ABDI
NIM. 1910817110008

Telah diperiksa dan terpenuhi semua persyaratan akademik, administrasi dan
disetujui untuk dipertahankan di hadapan dewan penguji

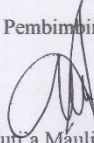
Banjarmasin, 31 Mei 2023

Pembimbing Utama,



Andreyan Rizky Baskara, S.Kom., M.Kom
NIP. 199307032019031011

Pembimbing Pendamping,



Muli'a Maulida, S.Kom., M.T.I
NIP. 198810272019032013

ABSTRAK

Indonesia menjadi salah satu negara dengan penggunaan media sosial tertinggi di era modern ini. Banyaknya jumlah pengguna media sosial di Indonesia menyebabkan banyaknya kasus ujaran kebencian yang terjadi di Indonesia. Kementerian Komunikasi dan Informatika Republik Indonesia telah menangani 3.640 kasus ujaran kebencian berbasis SARA di ruang digital sejak tahun 2018 hingga 26 April 2021. Keragaman demografis Indonesia sebagai negara kepulauan dengan 1.340 suku bangsa dan 718 bahasa daerah memperkaya gaya komunikasi pengguna media sosial dengan ekspresi budaya, logat, dan bahasa yang beragam. Namun, media sosial tidak memiliki aturan penggunaan bahasa yang digunakan, sehingga memungkinkan pengguna untuk menggunakan bahasa daerah. Hal ini menjadi tantangan tersendiri dalam mendeteksi ujaran kebencian karena terbatasnya sumber daya untuk bahasa daerah. Banyaknya ujaran kebencian yang menyebar di media sosial membutuhkan adanya deteksi ujaran kebencian dengan mengklasifikasikan unggahan yang mengandung konten tersebut. Klasifikasi ujaran kebencian secara manual membutuhkan biaya yang besar dan tidak akurat. Oleh karena itu, metode deteksi otomatis diperlukan untuk mengatasi masalah ini. *Ensemble method* adalah metode menggabungkan beberapa model kemudian menggabungkannya untuk menghasilkan akurasi yang lebih baik. *Ensemble method* dengan performa terbaik dalam penelitian ini menggunakan ekstraksi fitur N-gram dengan pengklasifikasi *Support Vector Machine*, *Naïve Bayes*, dan *Decision Tree* yang digabungkan menggunakan metode *soft voting* dengan data dibagi ke dalam rasio 80:20. *Ensemble method* mencapai akurasi 85,70%, presisi 85,99%, recall 85,70%, dan f1-score 85,64%.

Kata Kunci: Bahasa Banjar, Deteksi Ujaran Kebencian, Ensemble Method

ABSTRACT

Indonesia is one of the countries with the highest social media usage in this modern era. The large number of social media users in Indonesia has led to many cases of hate speech occurring in Indonesia. The Ministry of Communication and Information of the Republic of Indonesia has handled 3,640 cases of SARA-based hate speech in the digital space from 2018 to April 26, 2021. Indonesia's demographic diversity as an archipelago with 1,340 ethnic groups and 718 regional languages enriches the communication style of social media users with diverse cultural expressions, accents, and languages. However, social media does not have rules for language use, allowing users to use local languages. This is a challenge in detecting hate speech due to limited resources for local languages. The amount of hate speech spread on social media requires the detection of hate speech by classifying uploads containing such content. Manual classification of hate speech is costly and inaccurate. Therefore, automatic detection methods are needed to solve this problem. Ensemble method, is a method of combining several models and then combining them to produce better accuracy. The best-performing ensemble method in this study uses N-gram feature extraction with Support Vector Machine, Naïve Bayes, and Decision Tree classifiers combined using a soft voting method with the data divided into an 80:20 ratio. The ensemble method achieved 85.70% accuracy, 85.99% precision, 85.70% recall, and 85.64% f1-score.

Keywords: Banjar Language, Ensemble Method, Hate Speech Detection

LEMBAR PERSEMBAHAN

Penulis mempersembahkan Tugas Akhir kepada:

1. Ibu, Ayah, Kakak dan keluarga tercinta yang telah memberikan motivasi, dukungan moral dan materi, dan senantiasa mendoakan penulis akan keberlangsungan penyelesaian Tugas Akhir ini.
2. Ibu Dr. Ir. Yuslena Sari, S.Kom, M.Kom., selaku Koordinator Program Studi Teknologi Informasi yang selalu menyempatkan waktunya untuk memberikan bimbingan dan arahan kepada kami mahasiswa teknologi informasi untuk segera menyelesaikan Tugas Akhir ini.
3. Bapak Andreyan Rizky Baskara, S.Kom., M.Kom., selaku Dosen Pembimbing Utama yang selalu menyempatkan waktu untuk memberikan bimbingan, arahan, dan dukungan kepada penulis dari awal sampai akhir penyelesaian Tugas Akhir ini.
4. Ibu Muti'a Maulida, S.Kom., M.T.I selaku Dosen Pembimbing Pendamping dan Dosen Pembimbing Akademik yang senantiasa memberikan bimbingan, arahan, dan dukungan kepada penulis untuk segera menyelesaikan Tugas Akhir ini.
5. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dalam penyelesaian Laporan Tugas Akhir.
6. Eldy Yuda Kurniawan selaku kakak tingkat yang telah mendukung dan membantu dalam pengerjaan Tugas Akhir.
7. Teman-teman seperjuangan dari masa perkuliahan sampai tugas akhir yaitu, Digdo Aji Asrowi, Irvan Aulia Luthfi, Muhammad Afrizal Miqdad, Muhammad Ardhy Satrio Jati, Muhammad Arras, Muhammad Fachrurrazi, Muhammad Faidhorrahman, Muhammad Miftachul Falah dan seluruh Angkatan 2019 Program Studi Teknologi Informasi yang memotivasi dalam penyusunan Laporan Tugas Akhir.

KATA PENGANTAR

Puji dan syukur penulis ucapkan kehadiran Allah Subhanahu wa ta'ala, Tuhan Yang Maha Pengasih dan Penyayang yang telah memberikan kita berbagai macam nikmat dan rezeki, sehingga semua cita-cita serta harapan yang ingin kita capai menjadi lebih mudah dan bermanfaat untuk orang banyak. Sholawat dan salam tidak lupa penulis sampaikan kepada junjungan kita Nabi Besar Muhammad SAW yang telah membawa kita ke jalan yang terang benderang. Selain itu, atas limpahan rahmat serta karunia dari Allah SWT, sehingga penulis dapat menyelesaikan Tugas Akhir dengan Judul “Deteksi Ujaran Kebencian Bahasa Banjar Menggunakan *Ensemble Method*”.

Dalam penyusunan laporan ini, penulis menyadari sepenuhnya bahwa laporan ini tidak akan terwujud tanpa adanya bimbingan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Ibu, Ayah, Kakak dan keluarga tercinta yang telah memberikan motivasi, dukungan moral dan materi, dan senantiasa mendoakan penulis akan keberlangsungan penyelesaian Tugas Akhir ini.
2. Ibu Dr. Ir. Yuslena Sari, S.Kom, M.Kom., selaku Koordinator Program Studi Teknologi Informasi yang selalu menyempatkan waktunya untuk memberikan bimbingan dan arahan kepada kami mahasiswa teknologi informasi untuk segera menyelesaikan Tugas Akhir ini.
3. Bapak Andreyan Rizky Baskara, S.Kom., M.Kom., selaku Dosen Pembimbing Utama yang selalu menyempatkan waktu untuk memberikan bimbingan, arahan, dan dukungan kepada penulis dari awal sampai akhir penyelesaian Tugas Akhir ini.
4. Ibu Muti'a Maulida, S.Kom., M.T.I selaku Dosen Pembimbing Pendamping dan Dosen Pembimbing Akademik yang senantiasa memberikan bimbingan, arahan, dan dukungan kepada penulis untuk segera menyelesaikan Tugas Akhir ini.
5. Seluruh Dosen beserta Staf Program Studi Teknologi Informasi yang turut membantu dan mengarahkan dalam penyelesaian Laporan Tugas Akhir.

6. Eldy Yuda Kurniawan selaku kakak tingkat yang telah mendukung dan membantu dalam pengerjaan Tugas Akhir.
7. Teman-teman seperjuangan dari masa perkuliahan sampai tugas akhir yaitu, Digdo Aji Asrowi, Irvan Aulia Luthfi, Muhammad Afrizal Miqdad, Muhammad Ardhy Satrio Jati, Muhammad Arras, Muhammad Fachrurrazi, Muhammad Faidhorrahman, Muhammad Miftachul Falah dan seluruh Angkatan 2019 Program Studi Teknologi Informasi yang memotivasi dalam penyusunan Laporan Tugas Akhir.

Akhir kata, penulis menyampaikan terima kasih kepada semua pihak yang turut membantu dalam penyelesaian laporan Tugas Akhir ini. Penyusunan laporan Tugas Akhir ini disusun dengan sebaik-baiknya namun masih terdapat kekurangan di dalam penyusunan laporan Tugas Akhir ini, oleh karena itu saran dan kritik yang sifatnya membangun dari semua pihak sangat diharapkan, tidak lupa harapan penulis agar laporan ini dapat bermanfaat bagi para pembaca dan semua pihak yang membutuhkan.

Banjarmasin, Juni 2023
Penulis,

Muhammad Nur Abdi

DAFTAR ISI

SAMPUL DALAM.....	ii
LEMBAR PERNYATAAN.....	iii
LEMBAR PENGESAHAN.....	iv
LEMBAR PERSETUJUAN.....	v
ABSTRAK.....	vi
ABSTRACT.....	vii
LEMBAR PERSEMBAHAN.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR.....	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Landasan Teori.....	6
2.1.1 Ujaran Kebencian.....	6
2.1.2 Bahasa Banjar.....	6
2.1.3 <i>Ensemble Method</i>	7
2.1.4 <i>Natural Language Processing</i>	9
2.1.5 <i>Text Pre-Processing</i>	9
2.1.6 <i>Feature Extraction</i>	10
2.1.7 Klasifikasi.....	11
2.1.8 <i>Performance Evaluation Metrics</i>	13
2.2 Penelitian Terkait.....	14
2.2.1 <i>Detecting Hate Speech on Twitter Network Using Ensemble Machine Learning</i>	15

2.2.2	<i>Ensemble Method for Indonesian Twitter Hate Speech Detection....</i>	15
2.2.3	<i>Detecting Hate Speech Tweets and Abusive Tweets in Indonesian Language Using Random Forest and Support Vector Machine with Voting Classifier Technique</i>	16
2.2.4	<i>Hate Speech Spreader Detection using N-grams and Voting Classifier</i>	16
2.2.5	<i>Sentiment Analysis and Hate Speech Identification using Machine Learning, Deep Learning and Ensemble Models</i>	17
2.2.6	<i>Hate Speech and Offensive Content (HASOC) 2019: Hate Speech Detection.....</i>	17
2.2.7	<i>Twitter Sentiment Analysis Using an Ensemble Majority Vote Classifier.....</i>	18
2.2.8	<i>An Ensemble Approach for Classification and Prediction of Diabetes Mellitus Using Soft Voting Classifier</i>	18
2.2.9	<i>Abusive Language and Hate Speech Detection for Javanese and Sundanese Languages in Tweets: Dataset and Preliminary Study ...</i>	19
2.2.10	<i>Abusive Language and Hate Speech Detection for Indonesian-Local Language in Social MediaText.....</i>	20
2.2.11	<i>Voting Combinations Based Ensemble: A Hybrid Approach.....</i>	21
2.2.12	<i>Ensemble Learning Technique to Improve Classification Accuracy for Credit Data</i>	21
2.2.13	<i>Bagging Based Ensemble Classification Method on Imbalance Datasets</i>	22
2.3	Kerangka Pemikiran.....	29
BAB III METODOLOGI PENELITIAN		30
3.1	Alat dan Bahan Penelitian.....	30
3.1.1	Alat Penelitian	30
3.1.2	Bahan Penelitian	31
3.2	Alur Penelitian	31
3.2.1	Identifikasi Masalah	31
3.2.2	Studi Literatur.....	32
3.2.3	Pengumpulan Data.....	32
3.2.4	Melakukan Eksperimen	33
3.2.5	Evaluasi Performa	38
3.2.6	Implementasi Sistem	39
BAB IV HASIL DAN PEMBAHASAN		40

4.1	Pengumpulan Data	40
4.2	<i>Text Pre-Processing</i>	40
4.2.1	<i>Case Folding</i>	41
4.2.2	<i>Text Filtering</i>	42
4.3	<i>Translation</i> Bahasa Banjar	42
4.4	Pelabelan Data	43
4.5	Membangun Model <i>Ensemble Method</i>	44
4.5.1	<i>Import Library</i>	44
4.5.2	<i>Read Dataset</i>	46
4.5.3	<i>Ensemble Method</i>	46
4.5.4	<i>Train</i> dan Evaluasi Model	48
4.5.5	Validasi Model	49
4.6	Pengujian Model <i>Ensemble Method</i>	50
4.6.1	Pengujian <i>Base Model Ensemble Method</i>	50
4.6.2	Pengujian <i>Ensemble Method</i> Menggunakan Metode <i>Hard Voting</i> ...	52
4.6.3	Pengujian <i>Ensemble Method</i> Menggunakan Metode <i>Soft Voting</i>	53
4.7	Evaluasi Performa Model <i>Ensemble Method</i>	53
4.8	Analisis Hasil	56
4.9	Implementasi Model <i>Ensemble Method</i>	57
BAB V KESIMPULAN DAN SARAN		59
5.1	Kesimpulan	59
5.2	Saran	59
DAFTAR PUSTAKA		60
LAMPIRAN		65

DAFTAR TABEL

Tabel 2.1 Ringkasan Penelitian Terkait	23
Tabel 3.1 Alat Penelitian.....	30
Tabel 3.2 Contoh Data	33
Tabel 3.3 Proses <i>Case Folding</i>	34
Tabel 3.4 Proses <i>Punctuation Removal</i>	34
Tabel 3.5 Proses <i>Emoji Removal</i>	34
Tabel 3.6 <i>Translation</i> Bahasa Banjar.....	35
Tabel 4.1 Hasil <i>Case Folding</i>	42
Tabel 4.2 Hasil <i>Text Filtering</i>	42
Tabel 4.3 Kata yang Tidak Memiliki Padanan Kata pada Bahasa Banjar	43
Tabel 4.4 Hasil <i>Translation</i> bahasa Banjar.....	43
Tabel 4.5 Acuan Pemberian Label.....	44
Tabel 4.6 Accuracy dan Nilai Validasi	50
Tabel 4.7 Hasil Pengujian Menggunakan Metode <i>Hard Voting</i>	52
Tabel 4.8 Hasil Pengujian Menggunakan Metode <i>Soft Voting</i>	53
Tabel 4.9 Hasil Klasifikasi Data Validasi	54
Tabel 4.10 Data yang Berhasil Diklasifikasikan dengan Benar.....	54
Tabel 4.11 Data yang Tidak Berhasil Diklasifikasikan dengan Benar	54

DAFTAR GAMBAR

Gambar 1.1 Komentar Warga Kalimantan pada Postingan Sahbirin Noor	2
Gambar 2.2 <i>Bagging</i>	8
Gambar 2.3 <i>Hard Voting</i>	8
Gambar 2.4 <i>Soft Voting</i>	9
Gambar 2.5 <i>Decision Tree</i>	13
Gambar 2.6 Kerangka Pemikiran.....	29
Gambar 3.1 Alur Penelitian	31
Gambar 3.2 Tahapan Eksperimen	33
Gambar 3.3 Tahapan Pengujian	38
Gambar 3.4 Contoh Implementasi Sistem dengan Data Ujaran Kebencian	39
Gambar 3.5 Contoh Implementasi Sistem dengan Data Non-Ujaran Kebencian .	39
Gambar 4.1 Hasil Pengujian <i>Training Time</i>	52
Gambar 4.2 <i>Wordcloud</i> Ujaran Kebencian	55
Gambar 4.3 <i>Wordcloud</i> Non-Ujaran Kebencian.....	55
Gambar 4.4 Tampilan Sistem Deteksi Ujaran Kebencian bahasa Banjar.....	57
Gambar 4.6 Contoh Masukan yang Bukan Merupakan Ujaran Kebencian.....	58
Gambar 4.5 Contoh Masukan yang Merupakan Ujaran Kebencian	58