

**DETEKSI UJARAN KEBENCIAN DALAM BAHASA BANJAR  
MENGUNAKAN BI-LSTM DENGAN ATTENTION LAYER**

**SKRIPSI**

**Oleh:**

**LAILY RACHMAH**

**NIM. 2010817220013**



**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS LAMBUNG MANGKURAT**

**2024**

**DETEKSI UJARAN KEBENCIAN DALAM BAHASA BANJAR  
MENGUNAKAN BI-LSTM DENGAN ATTENTION LAYER**

**SKRIPSI**

Diajukan Untuk Memenuhi Salah Satu Syarat

Sarjana Strata-1 Teknologi Informasi

**Oleh:**

**LAILY RACHMAH**

**NIM. 2010817220013**



**PROGRAM STUDI TEKNOLOGI INFORMASI**

**FAKULTAS TEKNIK**

**UNIVERSITAS LAMBUNG MANGKURAT**

**2024**

## LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Laily Rachmah  
NIM : 2010817220013  
Fakultas : Teknik  
Prodi : Teknologi Informasi  
Judul Skripsi : Deteksi Ujaran Kebencian dalam Bahasa Banjar Menggunakan Bi-LSTM dengan Attention Layer  
Pembimbing Utama : Andreyan Rizky Baskara, S.Kom., M.Kom.

Dengan ini saya menyatakan bahwa dalam skripsi 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 tertulis diacu dalam naskah ini dan disebutkan dalam daftar rujukan.

Banjarmasin, Juni 2024

Laily Rachmah  
NIM. 2010817220013

## LEMBAR PENGESAHAN

**SKRIPSI PROGRAM STUDI S-1 TEKNOLOGI INFORMASI**  
**Deteksi Ujaran Kebencian dalam Bahasa Banjar Menggunakan Bi-LSTM**  
**dengan *Attention Layer***

Oleh

**Laily Rachmah (2010817220013)**

Telah dipertahankan di depan Tim Penguji pada 3 Juli 2024 dan dinyatakan

**LULUS**

**Komite Penguji :**

**Ketua :** Eka Setya Wijaya, S.T., M.Kom.  
NIP. 198205082008011010

**Anggota 1 :** Muhammad Fajrian Noor S.Kom., M.Kom.  
NIP. 199611092023211009

**Anggota 2 :** Muti'a Maulida, S.Kom., M.T.I.  
NIP. 198810272019032013

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



.....  
.....  
.....  
.....

**10 JUL 2024**

Banjarbaru, .....

Diketahui dan disahkan oleh:

**Wakil Dekan Bidang Akademik  
Fakultas Teknik ULM,**



**Koordinator Program Studi  
S-1 Teknologi Informasi,**



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

## ABSTRAK

Sarana media sosial, yang penting bagi individu dan kelompok dalam menyuarakan aspirasi, juga menjadi tempat meningkatnya ujaran kebencian karena sifat terbukanya. Informasi dapat dengan mudah diakses dan dikomentari, memicu kecenderungan ujaran kebencian. Pengguna bebas menggunakan bahasa daerah karena medsos tidak memiliki pedoman bahasa, dan keterbatasan sumber daya dalam memahami bahasa daerah menjadi kendala deteksi ujaran kebencian. Kompleksitas ini diperkuat oleh kebijakan platform media sosial seperti Instagram yang belum efektif dalam mendeteksi ujaran kebencian dalam bahasa daerah, seperti Bahasa Banjar. Kasus tersebut menunjukkan perlunya solusi lebih canggih, seperti otomatisasi deteksi ujaran kebencian dalam bahasa Banjar menggunakan deep learning. Bi-LSTM + Attention Layer adalah model jaringan saraf yang menggunakan Bidirectional Long Short Term Memory untuk memodelkan konteks sebelumnya dan berikutnya dalam teks, ditingkatkan dengan Attention layer yang menambahkan bobot pada data teks untuk meningkatkan performa klasifikasi ujaran kebencian. Model yang digunakan dalam penelitian ini banyak diterapkan karena mampu memberikan performa yang unggul dalam deteksi ujaran kebencian. Hal ini terbukti dari penelitian yang mengkombinasikan word embedding Word2Vec Skip-gram dengan model Bi-LSTM + Attention Layer, yang mencapai performa terbaik pada dataset ujaran kebencian bahasa Banjar normalisasi *lexical variation* dengan nilai akurasi 87.65%, presisi 87.78%, recall 87.61%, dan F1-score 87.62%.

Kata Kunci: Bahasa Banjar, Deteksi Ujaran Kebencian, Deep Learning, Bi-LSTM, Attention Layer

## ABSTRACT

*Social media platforms, which are essential for individuals and groups to express their aspirations and opinions, have also become a venue for the rise of hate speech due to their open nature. The ease with which information can be accessed and commented on by users triggers a tendency towards hate speech. Users are free to use regional languages as social media lacks guidelines on language use, and the limited resources for understanding regional languages pose a challenge in detecting hate speech. This complexity is compounded by the inadequacy of social media platform policies, such as Instagram, in effectively detecting hate speech in regional languages, as evidenced by cases of hate speech in the Banjar language. These cases highlight the need for more advanced solutions, such as automating the detection of hate speech in the Banjar language using deep learning. The Bi-LSTM + Attention Layer is a neural network model that employs Bidirectional Long Short Term Memory to model previous and subsequent contexts in text, enhanced with an Attention layer that adds weight to text data to improve hate speech classification performance. The model used in this research is widely applied due to its superior performance in hate speech detection. This is demonstrated by research combining Word2Vec Skip-gram word embedding with the Bi-LSTM + Attention Layer model, achieving the best performance on a dataset of normalized lexical variation in Bahasa Banjar hate speech with an accuracy of 87.65%, precision of 87.78%, recall of 87.61%, and an F1-score of 87.62%.*

*Keywords: Banjarese Language, Hate Speech Detection, Deep Learning, Bi-LSTM, Attention Layer*

## HALAMAN PERSEMBAHAN

Penulis mempersembahkan saya persembahkan kepada :

1. Ibu, Ayah dan Kakak-kakak tercinta yang selalu telah memberikan motivasi, dukungan moral dan materi, dan senantiasa mendoakan penulis akan keberlangsungan penyelesaian Skripsi ini.
2. Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Utama sekaligus Koordinator Program Studi Teknologi Informasi yang telah bersedia meluangkan waktunya untuk memberikan arahan, bimbingan dan dukungan kepada saya dengan penuh kesabaran agar skripsi ini dapat cepat terselesaikan.
3. Bapak Prof. Juhriyansyah Dalle, S.Pd., S.Si., M. Kom., Ph.D., selaku Dosen Pembimbing Akademik penulis, yang telah bersedia membantu dan memberikan bimbingan terbaik selama masa perkuliahan dengan penuh kesabaran.
4. Seluruh Dosen beserta Staf Administrasi Program Studi Teknologi Informasi yang turut mengarahkan dan membantu selama menyelesaikan skripsi.
5. Fajra Hanifa Nuridi Radam dan Kak Neysa Nisrina sebagai teman seperjuangan serta Kak Muhammad Nur Abdi yang telah membantu penulis serta memberikan motivasi, dukungan, kritik, dan saran selama masa perkuliahan hingga penyelesaian skripsi ini.
6. Seluruh teman-teman Angkatan 2020 Program Studi Teknologi Informasi serta kakak tingkat lainnya yang pernah membantu serta memberi semangat kepada penulis selama perkuliahan hingga penyelesaian skripsi ini.

## KATA PENGANTAR

Puji dan syukur tiada henti penulis ucapkan kehadirat 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 skripsi dengan Judul “Deteksi Ujaran Kebencian dalam Bahasa Banjar Menggunakan Bi-LSTM dengan *Attention Layer*”.

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. Allah Subhanahu Wa Ta'ala yang Maha Esa dan Maha Kuasa yang telah memberikan nikmat hidup, Kesehatan, dan jalan agar penulis dapat menyelesaikan skripsi ini dengan baik.
2. Ibu, Ayah, Kakak-kakak tercinta yang telah memberikan motivasi, dukungan moral dan materi, dan senantiasa mendoakan penulis akan keberlangsungan penyelesaian skripsi ini. Mama tercinta, sosok Wanita terhebat yang selalu mendo'akan, menyayangi, dan membuatkan makanan yang enak ketika pulang ke rumah. Ayah dan kakak laki-laki, sosok pria terhebat yang selalu merawat, menyayangi, dan memberikan dukungan material kepada penulis. Kakak perempuan, sosok terhebat yang selalu bisa mem-*back up* semua pekerjaan yang terlewatkan dan menemani penulis selama proses pengerjaan skripsi.
3. Bapak Andreyan Rizky Baskara S.Kom., M.Kom., selaku Dosen Pembimbing Utama sekaligus Koordinator Program Studi Teknologi Informasi yang telah bersedia meluangkan waktunya untuk memberikan arahan, bimbingan dan dukungan kepada saya dengan penuh kesabaran agar skripsi ini dapat cepat terselesaikan.



4. Teman seperjuangan saat perkuliahan dan skripsi yaitu Amazida, Annisa Mahfuzhah, Fajra Hanifa Nuridi Radam, Ghinaa Haniifah, Hilmiyah Amaliyah, Akbar, Fariz, Majdi, Oling, Pram, Shendy, dan Zaini, serta Kak Neysa Nisrina, Kak M. Yusuf Hasbullah dan Kak Muhammad Nur Abdi yang telah membantu penulis serta memberikan motivasi, dukungan, kritik dan saran kepada penulis selama masa perkuliahan hingga penyelesaian skripsi ini.

Penulis mengucapkan terimakasih kepada semua pihak yang turut berperan dalam menyelesaikan penyusunan skripsi ini. Meskipun skripsi ini telah disusun dengan baik berkat banyak bantuan, namun penulis menyadari bahwa skripsi ini masih banyak kekurangan. Oleh karena itu, penulis dengan tulus menerima saran dan kritikan yang membangun dari semua pihak. Atas perhatian dari pembaca, penulis ucapkan terimakasih.

Banjarmasin, Juni 2024

Laily Rachmah

NIM. 2010817220013

## DAFTAR ISI

HALAMAN SAMPUL LUAR .....	i
HALAMAN SAMPUL DALAM .....	ii
LEMBAR PERNYATAAN .....	iii
LEMBAR PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
HALAMAN PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI .....	x
DAFTAR TABEL .....	xiii
DAFTAR GAMBAR .....	xv
DAFTAR LAMPIRAN .....	xvi
DAFTAR RIWAYAT HIDUP .....	xvii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	5
1.4 Tujuan Penelitian .....	5
1.5 Manfaat Penelitian .....	5
BAB II TINJAUAN PUSTAKA .....	6
2.1 Landasan Teori .....	6
2.2.1 Bahasa Banjar .....	6
2.2.2 Ujaran Kebencian .....	6
2.2.3 <i>Natural Language Processing</i> .....	7
2.2.4 <i>Text Classification</i> .....	7
2.2.5 <i>Feature Extraction</i> .....	8
2.2.6 <i>Deep Learning</i> .....	8
2.2.7 <i>Word Embedding</i> .....	9
2.2.8 <i>Bidirectional Long Short Term Memory</i> .....	10
2.2.9 <i>Attention Mechanism</i> .....	13

2.2.10	<i>Text Pre-Processing</i> .....	14
2.2.11	<i>Performance Evaluation Metrics</i> .....	15
2.2	Penelitian Terkait.....	16
2.2.1	<i>Context-Aware Deep Learning Model for Detection of Roman Urdu Hate Speech on Social Media Platform</i> .....	16
2.2.2	<i>ETHOS: AN ONLINE HATE SPEECH DETECTION DATASET</i> ..	18
2.2.3	<i>Deep Learning for Detecting Cyberbullying Across Multiple Social Media Platforms</i> .....	18
2.2.4	<i>BiLSTM Model With Attention Mechanism for Sentiment Classification on Chinese Mixed Text Comments</i> .....	19
2.2.5	<i>BiCHAT: BiLSTM with deep CNN and Hierarchical Attention for Hate Speech Detection</i> .....	20
2.2.6	<i>Hate Speech Detection for Banjarese Languages on Instagram Using Machine Learning Methods</i> .....	20
2.2.7	<i>Abusive Language and Hate Speech Detection for Indonesian-Local Language in Social Media Text</i> .....	21
2.2.8	<i>Bidirectional Long Short Term Memory (LSTM) method and Word2vec extraction approach</i> .....	22
2.2.9	<i>Attention-Based BiLSTM For Negation Handling In Sentimen Analysis</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 .....	38

BAB IV HASIL DAN PEMBAHASAN .....	39
4.1 Dataset.....	39
4.1.1 Pengumpulan dan Pemilahan Data.....	39
4.1.2 <i>Pre-processing</i> Data.....	40
4.1.3 <i>Normalization of Lexical Variation</i> .....	43
4.1.4 Pelabelan Data.....	45
4.1.5 Penggabungan Data dan <i>Class Balancing</i> .....	45
4.2 Pengolahan dan Pembagian data.....	47
4.3 Membangun Model Bi-LSTM + Attention Layer.....	48
4.3.1 Import Library .....	48
4.3.2 Read Dataset.....	52
4.3.3 Word Embedding .....	52
4.3.4 Membuat Embedding Layer.....	53
4.3.5 Perbandingan Model Bi-LSTM dan Bi-LSTM + Attention Layer .....	54
4.3.6 Eksperimen Layer Bi-LSTM.....	55
4.3.7 Eksperimen <i>Hyperparameter</i> .....	57
4.3.8 Membuat Model Bi-LSTM + Attention Layer .....	59
4.4 Pengujian Model Bi-LSTM + Attention Layer pada Dataset Normalisasi dan Non-normalisasi .....	60
4.5 Evaluasi Performa Model.....	62
4.6 Analisis Hasil .....	63
4.7 Implementasi Model Bi-LSTM + Attention Layer .....	66
BAB V KESIMPULAN DAN SARAN.....	68
5.1 Kesimpulan .....	68
5.2 Saran.....	68
DAFTAR PUSTAKA .....	69
LAMPIRAN .....	74

## DAFTAR TABEL

Tabel 1.1 Komentar Miring pada Postingan Penerapan PSBB .....	2
Tabel 2.1 Ringkasan Penelitian Terkait .....	24
Tabel 3.1 Alat Penelitian.....	30
Tabel 3.2 Contoh Data Mentah .....	32
Tabel 3.3 Proses Lowercasing.....	34
Tabel 3.4 Proses Punctuation Removal.....	34
Tabel 3.5 Proses Emoji Removal .....	34
Tabel 3.6 Proses Tokenizing .....	35
Tabel 3.7 Proses Stop Word Removal .....	35
Tabel 4.1 Proses Pemilahan Data.....	40
Tabel 4.2 Hasil Lowercasing.....	42
Tabel 4.3 Hasil Text Filtering .....	42
Tabel 4.4 Hasil Stop-Word Removal .....	43
Tabel 4.5 Contoh Perbedaan Dialek Bahasa Banjar .....	43
Tabel 4.6 Contoh Perbedaan Dialek pada Fonem .....	43
Tabel 4.7 Proses Normalization of Lexical Variation.....	44
Tabel 4.8 Acuan Pemberian Label .....	45
Tabel 4.9 Perincian Data Training .....	46
Tabel 4.10 Hasil Evaluasi Model Imbalanced dan Balanced.....	47
Tabel 4.11 Nilai Val_Accuracy dan Val_Loss Embedding CBOW dan Skip-Gram .....	53
Tabel 4.12 Detail Arsitektur Model Bi-LSTM .....	54
Tabel 4. 13 Detail Arsitektur Model Bi-LSTM + Attention Layer.....	54
Tabel 4.14 Nilai Val_Accuracy dan Val_Loss Model Bi-LSTM dan Bi-LSTM + Attention Layer .....	55
Tabel 4.15 Rincian Eksperimen Jumlah Layer Bi-LSTM .....	56
Tabel 4.16 Nilai Val_Accuracy dan Val_Loss Eksperimen Layer Bi-LSTM .....	56
Tabel 4.17 Uji Coba Konfigurasi .....	57
Tabel 4.18 Percobaan Nilai Konfigurasi .....	58
Tabel 4.19 Pengaturan Konfigurasi Parameter Model Bi-LSTM + Attention Layer .....	61

Tabel 4.20 Hasil Pengujian pada Dataset Normalisasi dan Non-Normalisasi.....	61
Tabel 4.21 Hasil Pengujian pada Dataset Normalisasi menggunakan Confusion Matrix.....	62
Tabel 4.22 Hasil Pengujian pada Dataset Non-normalisasi menggunakan Confusion Matrix .....	63
Tabel 4.23 Hasil Rincian Prediksi Data .....	63
Tabel 4.24 Daftar Data Teks yang Berhasil Diprediksi dengan Benar .....	64
Tabel 4.25 Daftar Data Teks yang Berhasil Diprediksi dengan Salah.....	65
Tabel 4.26 Contoh Kata Ambigu dalam Bahasa Banjar .....	65
Tabel 4.27 Data yang Gagal Diubah pada Tahap Normalisasi .....	66

## DAFTAR GAMBAR

Gambar 2.1 Model LSTM.....	10
Gambar 2.2 Model Bi-LSTM.....	12
Gambar 2.3 Kerangka Pemikiran.....	29
Gambar 3.1 Alur Penelitian.....	31
Gambar 3.2 Tahapan Eksperimen.....	33
Gambar 3.3 Proses Normalisasi Lexical Variation Pada Bahasa Banjar.....	35
Gambar 3.4 Model Bi-LSTM + Attention Layer.....	37
Gambar 3.5 Contoh Implementasi Sistem.....	38
Gambar 4.1 Confusion Matrix Imbalanced Dataset.....	46
Gambar 4.2 Confusion Matrix Balanced Dataset.....	46
Gambar 4.3 Grafik Perbandingan Accuracy dan Loss Embedding CBOW dan Skip-Gram.....	52
Gambar 4.4 Grafik Accuracy dan Loss Perbandingan Bi-LSTM dan Bi-LSTM + Attention Layer.....	55
Gambar 4.5 Grafik Accuracy dan Loss pada Eksperimen Layer Bi-LSTM.....	56
Gambar 4.6 Grafik Pelatihan pada Train Set dari Jumlah Epoch.....	58
Gambar 4.7 Grafik Pelatihan pada Validation Set dari Jumlah Epoch.....	58
Gambar 4.8 Grafik Accuracy dan Loss pada Dataset Normalisasi.....	62
Gambar 4.9 Grafik Accuracy dan Loss pada Dataset Non-normalisasi.....	62
Gambar 4.10 Tampilan Sistem Deteksi Ujaran Kebencian Bahasa Banjar.....	66
Gambar 4.11 Contoh Masukan yang Merupakan Ujaran Kebencian.....	67
Gambar 4.12 Contoh Masukan yang Bukan Merupakan Ujaran Kebencian.....	67

## DAFTAR LAMPIRAN

Lampiran 1. Lembar Konsultasi Pembimbing .....	74
Lampiran 2. Kuesioner Penelitian Validasi Dataset .....	75
Lampiran 3. Corpus Lexical Variasi Bahasa Banjar .....	76