

# **Penerapan Metode *Long-Short Term Memory* dan *FastText* Pada Klasifikasi Ujaran Kebencian**

Diajukan Untuk Menyusun Skripsi  
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh :

Rachmaniah Kesuma Wardani

09021282126090

**Jurusan Teknik Informatika  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2024**

## LEMBAR PENGESAHAN SKRIPSI

### PENERAPAN METODE LONG-SHORT TERM MEMORY DAN FASTTEXT PADA KLASIFIKASI UJARAN KEBENCIAN

Oleh :

**Rachmaniah Kesuma Wardani**  
NIM: 09021282126090


Indralaya, 31 Desember 2024

Mengetahui,

Ketua Jurusan Teknik Informatika



Pembimbing,



**Desty Rodiah, S.Kom., M.T.**  
NIP. 198912212020122011

## TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Selasa tanggal 31 Desember 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Rachmaniah Kesuma Wardani  
Nim : 09021282126090  
Judul : Penerapan Metode *Long-Short Term Memory* dan *FastText* pada Klasifikasi Ujaran Kebencian

dan dinyatakan LULUS.

1. Ketua Penguji

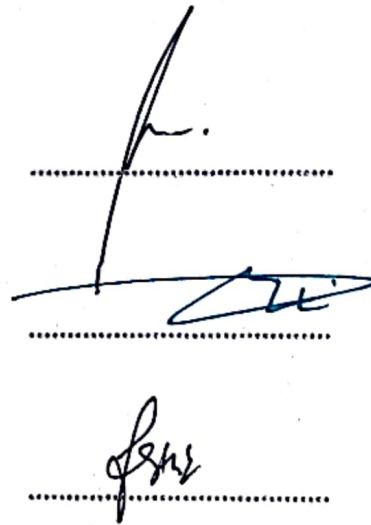
Rizki Kurniati, S.Kom., M.T.  
NIP. 199107122019032016

2. Penguji I

Osvari Arsalan, S.Kom., M.T.  
NIP. 198806282018031001

3. Pembimbing I

Desty Rodiah, S.Kom., M.T.  
NIP. 198912212020122011



Mengetahui,  
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, S.Kom, M.Sc, Ph.D.  
NIP. 198004182020121001

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Rachmaniah Kesuma Wardani  
NIM : 09021282126090  
Program Studi : Teknik Informatika  
Judul Skripsi : Penerapan Metode Long-Short Term Memory dan FastText pada Klasifikasi Ujaran Kebencian

**Hasil Pengecekan *Software iThenticate/Turnitin*: 20%**

Menyatakan bahwa laporan penelitian saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam penelitian ini maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 6 Januari 2024



Rachmaniah Kesuma Wardani

NIM. 09021282126090

## **MOTTO DAN PERSEMBAHAN**

"The path may be uncertain, but as long as you're walking, you're making progress."

Kupersembahkan karya tulis ini kepada:

- Kedua orang tua tersayang, Adik, dan Kakak
- Kucing-kucingku tersayang
- SVT, FCB, dan Tifosi
- Rekan seperjuangan Teknik Informatika

# APPLICATION OF THE *LONG-SHORT TERM MEMORY* METHOD AND *FASTTEXT* ON CLASSIFICATION OF HATE SPEECH

Oleh :

Rachmaniah Kesuma Wardani

09021282126090

## ABSTRACT

Hate speech on digital platforms poses a serious threat to social cohesion as it can lead to discrimination, conflict, and violence. This highlights the importance of developing automated detection systems to mitigate its impact and create safer digital spaces. To address this challenge, this study aims to develop a hate speech classification model in the Indonesian language using the Long Short-Term Memory (LSTM) method combined with the *Fasttext Word Embedding* approach. *Fasttext* was chosen for its ability to capture sub-word meanings and handle spelling variations, while LSTM excels in processing sequential data and understanding textual context comprehensively. The dataset used in this study comprises 2,980 social media comments categorized into two classes: hate speech and non-hate speech. The best-performing model, configured with 128 neurons in the LSTM layer, a *dropout rate* of 0.7, a *learning rate* of 0.1, a *batch size* of 64, and trained for 30 *epochs*, achieved an accuracy of 92.24%, a *precision* of 92.58%, a *recall* of 92.58%, and an *F1-score* of 92.58%. These results demonstrate that the developed model exhibits strong performance in classifying comments with high accuracy and robust generalization capabilities.

**Keywords:** Hate Speech, Classification, LSTM, *Fasttext*

**PENERAPAN METODE *LONG-SHORT TERM MEMORY* DAN *FASTTEXT*  
PADA  
KLASIFIKASI UJARAN KEBENCIAN**

Oleh :

Rachmaniah Kesuma Wardani

09021282126090

**ABSTRAK**

Ujaran kebencian di platform digital menjadi ancaman serius bagi kohesi sosial karena dapat memicu diskriminasi, konflik, dan kekerasan. Kondisi ini menekankan pentingnya pengembangan sistem deteksi otomatis untuk mengurangi dampaknya serta menciptakan ruang digital yang lebih aman. Untuk menjawab tantangan ini, penelitian ini bertujuan untuk mengembangkan model klasifikasi ujaran kebencian dalam bahasa Indonesia dengan menggunakan metode *Long Short-Term Memory* (LSTM) yang dikombinasikan dengan pendekatan *Word Embedding Fasttext*. *Fasttext* dipilih karena kemampuannya menangkap makna sub-kata dan mengatasi variasi ejaan, sementara LSTM unggul dalam menangani urutan data dan konteks teks secara mendalam. Dataset yang digunakan dalam penelitian ini terdiri dari 2.980 data komentar media sosial yang terbagi menjadi dua kelas: ujaran kebencian dan non-ujaran kebencian. Model terbaik, dengan parameter 128 *neuron* pada LSTM, *dropout rate* sebesar 0.7, *learning rate* sebesar 0.1, *batch size* 64, dan pelatihan selama 30 *epoch*, berhasil mencapai akurasi sebesar 92.24%, *precision* sebesar 92.58%, *recall* sebesar 92.58%, dan *F1-score* sebesar 92.58%, menunjukkan bahwa model yang dikembangkan memiliki performa yang kuat dalam mengklasifikasikan komentar dengan akurasi tinggi dan kemampuan generalisasi yang baik.

**Kata Kunci:** Ujaran Kebencian, Klasifikasi, LSTM, *Fasttext*

## KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penerapan Metode *Long-short term memory* dan *Fasttext* pada Klasifikasi Ujaran Kebencian". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Strata 1 pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa keberhasilan ini tidak terlepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua tersayang, yang telah mendukung, membimbing, dan mendoakan dan memberikan sehingga penulis dapat menyelesaikan studi dengan baik, khususnya dalam menyelesaikan skripsi ini.
2. Ibu Desty Rodiah S.Kom., M.T selaku dosen pembimbing yang telah dengan sabar membimbing penulis selama penyusunan skripsi ini, memberikan arahan dan masukan yang sangat berharga bagi penulis.
3. Dosen-dosen dan staf Program Studi Teknik Informatika, yang telah memberikan ilmu dan pengetahuannya kepada penulis selama berkuliah di Universitas Sriwijaya.
4. Sahabat-sahabat yang selalu memberikan dukungan dan semangat kepada penulis selama proses penyusunan skripsi ini.



5. Teman-teman jurusan Teknik Informatika yang tidak dapat dituliskan satu persatu yang telah memberikan bantuan, motivasi dan semangat serta menemani selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis menerima kritik dan saran yang membangun dari semua pihak demi perbaikan skripsi ini di masa depan. Semoga skripsi ini bermanfaat bagi penulis, keluarga, almamater, dan masyarakat luas, khususnya dalam upaya memerangi ujaran kebencian di dunia digital.

Palembang, 23 Desember 2024

Penyusun,

Rachmaniah Kesuma Wardani

NIM. 09021282126090

## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	ii
TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI .....	iii
SURAT KETERANGAN PENGECEKAN .....	iv
SIMILARITY .....	iv
MOTTO DAN PERSEMBAHAN .....	v
ABSTRACT .....	vi
ABSTRAK .....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI .....	x
BAB I PENDAHULUAN .....	I-1
<u>1.1</u> Pendahuluan .....	I-1
<u>1.2</u> Latar Belakang.....	I-1
<u>1.3</u> Rumusan Masalah .....	I-5
<u>1.4</u> Tujuan Penelitian.....	I-5
<u>1.5</u> Manfaat Penelitian.....	I-5
<u>1.6</u> Batasan Masalah.....	I-6
<u>1.7</u> Sistematika Penulisan.....	I-6
<u>1.8</u> Kesimpulan.....	I-8
<b>BAB II KAJIAN LITERATUR.....</b>	<b>II-1</b>
2.1 Pendahuluan .....	II-1
2.2 Landasan Teori .....	II-1
2.2.1 Hate speech .....	II-1
2.2.2 Klasifikasi .....	II-2
2.2.3 Natural Language Processing.....	II-2
2.2.4 <i>Word Embedding</i> .....	II-5
2.2.5 <i>Fasttext</i> .....	II-6
2.2.6 <i>Long-short term memory</i> .....	II-7
2.2.7 Confusion matrix.....	II-11
2.2.8 Rational Unified Process.....	II-12
2.3 Penelitian Lain yang Relevan.....	II-13

2.4 Kesimpulan.....	II-16
<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>III-1</b>
3.1 Pendahuluan .....	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis Data .....	III-1
3.2.2 Sumber Data.....	III-2
3.2.3 Metode Pengumpulan Data.....	III-2
3.3 Tahapan Penelitian.....	III-3
3.3.1 Kerangka Kerja .....	III-4
3.3.2 Kriteria Pengujian .....	III-8
3.3.3 Format Data Pengujian.....	III-8
3.3.4 Alat yang Digunakan dalam Penelitian .....	III-9
3.3.5 Pengujian Penelitian.....	III-10
3.3.6 Analisis Hasil Pengujian dan Kesimpulan .....	III-11
3.4 Metode Pengembangan Perangkat Lunak .....	III-11
3.4.1 Fase Insepsi .....	III-11
3.4.2 Fase Elaborasi .....	III-11
3.4.3 Fase Konstruksi.....	III-11
3.4.4 Fase Transisi.....	III-11
3.5 Manajemen Proyek Penelitian.....	III-12
3.6 Kesimpulan.....	III-17
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>	<b>IV-1</b>
4.1 Pendahuluan .....	IV-1
4.2 Fase Insepsi .....	IV-1
4.2.2 Kebutuhan Sistem .....	IV-2
4.2.3 Analisis dan Desain.....	IV-4
4.2.3.1 Analisis Kebutuhan Perangkat Lunak.....	IV-5
4.2.3.2 Analisis Data.....	IV-5
4.2.3.3 Analisis Pre-Processing.....	IV-5
4.2.3.4 Analisis Proses Klasifikasi.....	IV-18
4.2.4 Implementasi .....	IV-18
4.2.4.1 Use Case .....	IV-19

4.2.4.2	Tabel Definisi Aktor.....	IV-20
4.2.4.3	Tabel Definisi Use Case.....	IV-20
4.2.4.4	Tabel Skenario Use Case .....	IV-21
4.3	Fase Elaborasi.....	IV-27
4.3.1	Perancangan Data.....	IV-27
4.3.2	Rancangan Interface.....	IV-28
4.3.3	Kebutuhan Sistem .....	IV-29
4.3.4	Diagram Activity.....	IV-30
4.3.5	<i>Sequence</i> Diagram.....	IV-34
4.4	Fase Konstruksi .....	IV-36
4.4.1	Kebutuhan Sistem .....	IV-36
4.4.2	Diagram Kelas.....	IV-37
4.4.3	Implementasi .....	IV-38
4.4.3.1	Implementasi Kelas.....	IV-38
4.4.3.2	Implementasi Antarmuka .....	IV-39
4.5	Fase Transisi .....	IV-40
4.5.1	Pemodelan Bisnis .....	IV-41
4.5.2	Rencana Pengujian .....	IV-41
4.5.3	Implementasi .....	IV-42
4.5.3.1	Pengujian Use Case Preprocessing Dataset .....	IV-43
4.5.3.2	Pengujian <i>Use Case Training</i> model LSTM.....	IV-44
4.5.3.3	Pengujian <i>Use Case</i> Klasifikasi Data.....	IV-45
4.6	Kesimpulan.....	IV-46
<b>BAB V HASIL DAN ANALISIS PENELITIAN .....</b>		<b>V-1</b>
5.1	Pendahuluan .....	V-1
5.2	Data Hasil Penelitian .....	V-1
5.2.1	Konfigurasi percobaan .....	V-1
5.2.2	Data Hasil Konfigurasi.....	V-7
5.2.2.1	Hasil Konfigurasi Skenario 1.....	V-7
5.2.2.2	Hasil Konfigurasi 2.....	V-9
5.2.2.3	Hasil Konfigurasi 3.....	V-10
5.2.2.4	Hasil Konfigurasi 4.....	V-12

5.2.2.5 Hasil Konfigurasi 5.....	V-13
5.2.3 Analisis Hasil Penelitian .....	V-15
5.3 Kesimpulan.....	V-18
<b><u>BAB VI KESIMPULAN DAN SARAN</u></b> .....	VI-1
6.1 Pendahuluan .....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2
<b>DAFTAR PUSTAKA</b> .....	xvii
<b><u>LAMPIRAN</u></b> .....	xx

## DAFTAR TABEL

Tabel II-1. <i>Confusion Matrix</i> .....	III-11
Tabel III-1. Jumlah Kategori Data .....	III-2
Tabel III-2. Pembagian Dataset.....	III-2
Tabel III-3. Tabel Rancangan <i>Confusion matrix</i> Hasil .....	III-9
Tabel III-4. Tabel Rancangan Hasil Pengujian .....	III-9
Tabel III-5. Perencanaan Aktivitas Penelitian .....	III-13
Tabel IV-1. Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Kebutuhan Non-Fungsional.....	IV-4
Tabel IV-3. Jumlah Kategori Data .....	IV-5
Tabel IV-4. Contoh Dataset .....	IV-6
Tabel IV-5. Proses <i>Cleaning Data</i> .....	IV-7
Tabel IV-6. Hasil Proses <i>Case Folding</i> .....	IV-9
Tabel IV-7. Hasil Proses <i>Tokenizing</i> .....	IV-10
Tabel IV-8. Hasil Proses <i>Stopword Removal</i> .....	IV-12
Tabel IV-9. Hasil Proses <i>Stemming</i> .....	IV-13
Tabel IV-10. Contoh <i>Word Vector</i> 300 Dimensi .....	IV-14
Tabel IV-11. Contoh <i>Embedding</i> .....	IV-18
Tabel IV-12. Contoh bobot dan bias .....	IV-24
Tabel IV-13. Tabel Definisi Aktor.....	IV-20
Tabel IV-14. Tabel Definisi <i>Use Case</i> .....	IV-20
Tabel IV-15. Tabel Skenario <i>Preprocessing</i> data .....	IV-21
Tabel IV-16. Tabel Skenario <i>Training</i> model LSTM .....	IV-24
Tabel IV-17. Tabel Skenario Klasifikasi <i>Hate speech</i> .....	IV-26
Tabel IV-18. Tabel Implementasi Kelas .....	IV-38
Tabel IV-19. Tabel Rencana Pengujian <i>Preprocessing</i> Dataset .....	IV-41
Tabel IV-20. Tabel Rencana Pengujian <i>Training</i> model LSTM.....	IV-42
Tabel IV-21. Tabel Rencana Pengujian Klasifikasi .....	IV-42
Tabel IV-22. Tabel Hasil Pengujian <i>use case Preprocessing</i> Dataset .....	IV-43
Tabel IV-23. Tabel Hasil Pengujian <i>use case Training</i> model LSTM .....	IV-44
Tabel IV-24. Tabel Hasil Pengujian <i>use case</i> Klasifikasi Data .....	IV-45
Tabel V-1. Konfigurasi Parameter .....	V-2
Tabel V-2. Konfigurasi Skenario 1 .....	V-5
Tabel V-3. Konfigurasi Skenario 2 .....	V-5
Tabel V-4. Konfigurasi Skenario 3 .....	V-6
Tabel V-5. Konfigurasi Skenario 4 .....	V-7
Tabel V-6. Konfigurasi Skenario 5 .....	V-7
Tabel V-7. Hasil Konfigurasi Skenario 1 .....	V-9
Tabel V-8. Hasil Konfigurasi Skenario 2.....	V-10
Tabel V-9. Hasil Konfigurasi Skenario 3.....	V-12
Tabel V-10. Hasil Konfigurasi Skenario 4.....	V-13

Tabel IV-2. Hasil Konfigurasi Skenario 5 ..... V-14

## DAFTAR GAMBAR

Gambar II-1. Tahapan <i>Cleaning</i> .....	II-4
Gambar II-2. Tahapan <i>Case Folding</i> .....	II-4
Gambar II-3. Tahapan <i>Tokenize</i> .....	II-5
Gambar II-4. Tahapan <i>Stemming</i> .....	II-5
Gambar II-5. Tahapan <i>Stop Words Removal</i> .....	II-6
Gambar II-6. Struktur LSTM .....	II-8
Gambar III-1. Alur tahapan penelitian .....	III-4
Gambar III-2. Diagram Alur Kerangka Kerja.....	III-5
Gambar IV-1. Diagram <i>Use Case</i> .....	IV-19
Gambar IV-2. Rancangan Halaman <i>Preprocessing</i> .....	IV-28
Gambar IV-3. Rancangan Halaman <i>Training</i> .....	IV-29
Gambar IV-4. Rancangan Halaman Klasifikasi.....	IV-29
Gambar IV-5. Diagram <i>Activity Preprocessing</i> Dataset.....	IV-31
Gambar IV-6. Diagram <i>Activity Training</i> model LSTM .....	IV-32
Gambar IV-7. Diagram <i>Activity</i> Klasifikasi Teks.....	IV-33
Gambar IV-8. Diagram <i>Sequence Preprocessing</i> Dataset.....	IV-34
Gambar IV-9. Diagram <i>Sequence Training</i> LSTM.....	IV-35
Gambar IV-10. Diagram <i>Sequence</i> Klasifikasi .....	IV-36
Gambar IV-11. Diagram Kelas .....	IV-37
Gambar IV-12. Tampilan antarmuka <i>Preprocessing</i> Dataset.....	IV-39
Gambar IV-13. Tampilan antarmuka <i>Training</i> model LSTM .....	IV-40
Gambar IV-14. Tampilan antarmuka Klasifikasi Data .....	IV-40
Gambar V-1. Perbandingan <i>Accuraccy</i> Model .....	V-15



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Pada bab pendahuluan akan dibahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, kesimpulan serta gambaran umum dari penelitian yang dilakukan penulis. Pada bab ini juga akan dijelaskan metode yang akan digunakan dalam penelitian dan penelitian yang terkait yang telah diteliti sebelumnya.

### **1.2 Latar Belakang**

Ujaran kebencian bagaikan virus yang marak di berbagai platform daring, menghadirkan dampak negatif yang sangat serius seperti diskriminasi, kekerasan, dan polarisasi sosial. Fenomena ini tidak hanya merusak hubungan antarindividu, tetapi juga dapat mengganggu kohesi sosial dalam masyarakat (Benesch, 2014). Ujaran kebencian yang tersebar secara masif dan cepat melalui internet memperkuat stereotip, memicu permusuhan antar kelompok, dan bahkan dapat berujung pada tindakan kekerasan fisik (Citron, 2014). Oleh karena itu, klasifikasi ujaran kebencian ini dapat berguna bagi pembaca untuk dapat meningkatkan kesadaran publik terhadap dampak negatif dari ujaran kebencian. Salah satu metode yang dapat digunakan untuk tujuan ini adalah Long Short-Term Memory (LSTM).

Metode Long Short-Term Memory (LSTM) mempunyai kemampuan yang unggul dalam menangani data urutan panjang dan mengingat informasi dalam jangka waktu yang lama (Hochreiter & Schmidhuber, 1997). LSTM merupakan modifikasi dari RNN (*Recurrent Neural Network*) yang dirancang khusus untuk mengatasi masalah *vanishing gradient* yang sering muncul pada jaringan berulang, terutama ketika harus memproses urutan data yang panjang (Bengio et al., 1994). Kelebihan utama LSTM terletak pada arsitekturnya yang mampu mempertahankan informasi penting dari urutan data yang panjang serta kemampuan untuk mengingat dan melupakan informasi secara selektif melalui mekanisme yang disebut *gates* (Gers et al., 2000). Hal ini sangat relevan dalam pemrosesan bahasa alami (NLP), di mana konteks kata-kata dalam kalimat dan paragraf memainkan peran krusial dalam menentukan makna keseluruhan (Goldberg, 2017).

Pada proses klasifikasi berbasis LSTM, penggunaan *Word Embedding* menjadi sangat penting untuk meningkatkan akurasi dan efisiensi sistem (Mikolov et al., 2013). *Word Embedding* memungkinkan representasi kata-kata dalam bentuk vektor yang dapat diproses oleh algoritma *machine learning*, sehingga konteks dan hubungan antar kata dapat dipahami dengan lebih baik (Pennington et al., 2014). *Fasttext* adalah salah satu teknik *Word Embedding* yang memiliki kelebihan dalam mempertimbangkan sub-kata. Berbeda dengan teknik *Word Embedding* lainnya seperti *Word2Vec* yang hanya menghasilkan satu vektor per kata, *Fasttext* memperlakukan setiap

kata sebagai kumpulan *n-gram* karakter, yang memungkinkan model untuk menangkap makna kata-kata yang jarang muncul atau tidak dikenal dengan lebih baik (Bojanowski et al., 2016). Pendekatan ini juga meningkatkan kemampuan model dalam memahami variasi ejaan dan bentuk kata yang berbeda, yang penting dalam pemrosesan bahasa alami (NLP) (Mikolov et al., 2017).

Kelebihan *Fasttext* juga termasuk kemampuannya untuk menghasilkan representasi kata yang robust terhadap variasi dan kesalahan ejaan, yang sering kali menjadi tantangan dalam pemrosesan teks informal seperti komentar di media sosial (Bojanowski et al., 2016). Hal ini semakin memperkuat kombinasi *Fasttext* dan LSTM dalam menghadapi tantangan yang ada dalam klasifikasi ujaran kebencian. Dengan menggunakan *Fasttext* untuk menghasilkan *embedding* kata yang kuat dan menggunakan LSTM untuk mengolah urutan kata-kata tersebut, sistem yang dihasilkan diharapkan dapat mencapai tingkat kinerja yang tinggi dalam mendeteksi ujaran kebencian, sebagaimana diukur melalui metrik seperti *Precision*, *Recall*, dan *F1-score* (Sandesara et al., 2020).

Kombinasi metode LSTM dan *Fasttext* ini mungkin dapat meningkatkan kinerja sistem klasifikasi ujaran kebencian. *Fasttext*, dengan kemampuannya menghasilkan representasi kata yang lebih rumit dan menangkap makna morfologis dari kata-kata, menyediakan input yang lebih informatif dan kontekstual bagi LSTM (Bojanowski et al., 2016). Input yang

lebih rumit ini memungkinkan LSTM untuk lebih efektif dalam memproses urutan teks dan mempertahankan informasi penting dari konteks kata-kata tersebut (Hochreiter & Schmidhuber, 1997). Dengan kata lain, *Fasttext* membantu LSTM untuk memahami teks dengan lebih baik, sementara LSTM memanfaatkan informasi tersebut untuk membuat prediksi yang lebih akurat mengenai adanya ujaran kebencian (Mikolov et al., 2017).

Dalam penelitian sebelumnya yang berjudul “Penerapan Metode *Long Short Term Memory* Untuk Klasifikasi Pada *Hate speech*” menggunakan data sekunder yang berasal dari *Kaggle* yaitu *Indonesian Abusive and Hate speech*. Data ini memiliki 13.169 baris dengan 13 atribut dengan nilai 0 untuk nilai ‘tidak’ dan 1 untuk nilai ‘ya’ berhasil mendapatkan nilai akurasi pada data latih sebesar 86.23% dan akurasi pada data validasi sebesar 87.10% dengan *epoch* sebanyak 10. (Kholifatullah & Prihanto, 2023).

Secara keseluruhan, penelitian ini bertujuan untuk mengembangkan sistem klasifikasi ujaran kebencian yang lebih akurat dan andal dengan menggabungkan kekuatan metode LSTM dan *Fasttext*. Kombinasi ini diharapkan tidak hanya dapat meningkatkan akurasi deteksi tetapi juga dapat menyediakan dasar yang kuat bagi pengembangan metode klasifikasi teks lainnya di masa depan. Dengan demikian, hasil penelitian ini tidak hanya memberikan kontribusi pada bidang NLP tetapi juga berkontribusi pada upaya menciptakan lingkungan digital yang lebih aman dan positif.

### 1.3 Rumusan Masalah

Berdasarkan permasalahan yang telah diuraikan pada latar belakang maka rumusan masalah yang dapat disimpulkan pada penelitian ini adalah sebagai berikut :

1. Bagaimana penerapan *Fasttext* dan LSTM pada klasifikasi ujaran kebencian?
2. Bagaimana tingkat kinerja klasifikasi ujaran kebencian dengan menggunakan metode *Fasttext* dan LSTM yang berupa *Accuraccy*, *Precision*, *Recall*, dan *F1-score*?

### 1.4 Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini adalah sebagai berikut :

1. Mengetahui langkah penerapan *Fasttext* dan LSTM pada klasifikasi ujaran kebencian.
2. Mengetahui tingkat kinerja klasifikasi ujaran kebencian dengan menggabungkan metode *Fasttext* dan LSTM yang berupa *Accuraccy Precision*, *Recall*, dan *F1-score*.

### 1.5 Manfaat Penelitian

Adapun manfaat pada penelitian ini adalah sebagai berikut :

1. Memperbaiki efektivitas deteksi pada teks informal, seperti komentar di media sosial, yang sering mengandung variasi bahasa.
2. Memperluas pemahaman tentang bagaimana metode LSTM dan *Fasttext* dapat memproses dan memahami konteks teks dengan lebih baik.

3. Hasil penelitian ini dapat menjadi referensi untuk penelitian selanjutnya.

### **1.6 Batasan Masalah**

Adapun batasan masalah pada penelitian ini adalah sebagai berikut :

1. Klasifikasi terdiri dari 2 label yaitu, *Hate-Speech* dan *Non-Hate-Speech*
2. Data yang digunakan dalam penelitian ini dalam Bahasa Indonesia.

### **1.7 Sistematika Penulisan**

Adapun sistematika penulisan pada penelitian ini adalah sebagai berikut :

#### **BAB I. PENDAHULUAN**

Bab ini menguraikan mengenai latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan pada penelitian ini.

#### **BAB II. KAJIAN LITERATUR**

Bab ini membahas mengenai dasar-dasar teori dan ini juga menguraikan penelitian-penelitian terdahulu yang terkait dengan penelitian ini.

#### **BAB III. METODOLOGI PENELITIAN**

Bab ini berisi pembahasan mengenai metodologi dan tahapan perancangan penelitian seperti pengumpulan data, metode pengembangan perangkat lunak, dan manajemen proyek penelitian.

#### **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab IV akan dijelaskan perancangan, implementasi, rencana pengujian serta penerapan metode *Long-short term memory* dan *Fasttext* dalam sistem klasifikasi ujaran kebencian.

#### **BAB V. HASIL DAN ANALISIS PENELITIAN**

Pada bab V akan membahas hasil penelitian dan analisis sistem klasifikasi ujaran kebencian yang telah dibuat.

#### **BAB VI. KESIMPULAN DAN SARAN**

Pada bab VI diisi dengan kesimpulan dari analisis pada bab-bab sebelumnya serta saran yang dapat diberikan penulis untuk pengembangan lebih lanjut mengenai penerapan metode *Long-short term memory* dan *Fasttext* pada sistem klasifikasi ujaran kebencian.

### **1.8 Kesimpulan**

Berdasarkan latar belakang penelitian yang telah diuraikan di atas, penelitian ini akan mengklasifikasi teks yang termasuk ke dalam jenis-jenis *Hate speech* dengan menggunakan metode Long Short Term Memory (LSTM) dan *Fasttext*

## DAFTAR PUSTAKA

- Alkhairi, P., Perdana Windarto, A., & Masjun Efendi, M. (2024). Optimasi LSTM Mengurangi Overfitting untuk Klasifikasi Teks Menggunakan Kumpulan Data Ulasan Film Kaggle IMDB. *Technology and Science (BITS)*, 6(2). <https://doi.org/10.47065/bits.v6i2.5850>
- Badan Pengembangan dan Pembinaan Bahasa. (2021). *Kamus Besar Bahasa Indonesia (KBBI)*.
- Benesch, S. (2014). *COUNTERING DANGEROUS SPEECH: NEW IDEAS FOR GENOCIDE PREVENTION*.
- Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 2, 157–166. <https://doi.org/doi:10.1109/72.279181>
- Boehm, B. W. (1988). *A Spiral Model of Software Development and Enhancement*. <https://doi.org/10.1109/2.59>
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2016). *Enriching Word Vectors with Subword Information*.
- Citron, D. K. (2014). *Hate Crimes in Cyberspace*. Harvard University Press.
- Fajri, F. N., & Syaiful, S. (2022). Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2635>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing* (Vol. 7). Springer International Publishing. <https://doi.org/10.1007/978-3-031-02165-7>
- Hastuti, R. P., Riona, V., & Hardiyanti, M. (2023). Content Retrieval dengan *Fasttext Word Embedding* pada Learning Management System Olimpiade. *Journal of Internet and Software Engineering (JISE)*, 4(1), 18. <https://doi.org/https://doi.org/10.22146/jise.v4i1.6766>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). *Bag of Tricks for Efficient Text Classification*. <http://arxiv.org/abs/1607.01759>



- Kamal Wisyaldin, M., Maya Luciana, G., Pariaman, H., & Pembangunan Jawa Bali, P. (2020). *Pendekatan Long Short-Term Memory untuk Memprediksi Kondisi Motor 10 kV pada PLTU Batubara*. 9(2). <https://doi.org/10.33322/kilat.v9i2.997>
- Khairunisa, N. K., & Hendikawati, P. (2024). Long Short-Term Memory and Gated Recurrent Unit Modeling for Stock Price Forecasting. *Jurnal Matematika, Statistika Dan Komputasi*, 21(1), 321–333. <https://doi.org/10.20956/j.v21i1.35930>
- Kholifatullah, B. A. H., & Prihanto, A. (2023). Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech. *Journal of Informatics and Computer Science (JINACS)*, 292–297. <https://doi.org/10.26740/jinacs.v4n03.p292-297>
- Kruchten, P. (1999). *What Is the Rational Unified Process?* <https://www.researchgate.net/publication/220018244>
- Kruchten, P. (2001). *The Rational Unified Process-An Introduction*. <https://www.researchgate.net/publication/220018149>
- Marwan, M. A., Umniati, N., Tjiptanata, R. A., & Budiyarto, R. (2022). PENERAPAN METODE RATIONAL UNIFIED PROCESS (RUP) DALAM PEMBUATAN WEB PEMBELAJARAN ELEKTRONIK UNTUK SEKOLAH MENENGAH PERTAMA. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 7(2), 137–146. <https://doi.org/10.36341/rabit.v7i2.2457>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. <http://arxiv.org/abs/1301.3781>
- Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2017). *Advances in Pre-Training Distributed Word Representations*.
- Nurdin, A., Anggo, B., Aji, S., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA *WORD EMBEDDING* WORD2VEC, GLOVE, DAN *FASTTEXT* PADA KLASIFIKASI TEKS. *Jurnal TEKNOKOMPAK*, 14(2), 74.
- Oruh, J., Viriri, S., & Adegun, A. (2022). Long Short-Term Memory Recurrent Neural Network for Automatic Speech Recognition. *IEEE Access*, 10, 30069–30079. <https://doi.org/10.1109/ACCESS.2022.3159339>
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. <https://doi.org/10.3115/v1/D14-1162>
- Putra, J. A., & Budi, A. (2020). PENERAPAN NATURAL LANGUAGE PROCESSING DALAM APLIKASI CHATBOT SEBAGAI MEDIA PENCARIAN INFORMASI DENGAN MENGGUNAKAN REACT (STUDI KASUS: INSTITUT BISNIS DAN INFORMATIKA KWIK KIAN GIE). *Jurnal Informatika Dan Bisnis*, 9(2).

- Putra, R. P., & Karo, K. (2022). *HATE SPEECH: DELINQUENCY OF INDONESIA ITE LAW FREE SPEECH AND LAW IOF INFORMATION AND ELECTRONIC TRANSACTIONS*.  
[https://scholar.google.co.id/citations?view\\_op=view\\_citation&hl=en&user=ZlWA-NegAAAAJ&citation\\_for\\_view=ZlWANegAAAAJ:CQrscp367IIC](https://scholar.google.co.id/citations?view_op=view_citation&hl=en&user=ZlWA-NegAAAAJ&citation_for_view=ZlWANegAAAAJ:CQrscp367IIC)
- Sandesara, A., Parikh, S., Sapovadiya, P., & Rahevar, M. (2020). A Comparative Study On Speech Emotion Recognition. *International Journal of Research in Engineering, Science and Management*, 3(11), 25–35.  
<https://doi.org/10.47607/ijresm.2020.366>
- Widhiyasana, Y., Semiawan, T., Gibran, I., Mudzakir, A., & Noor, M. R. (2021). Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification). In *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 10, Issue 4).
- Wijayanto, I. R., Cholissodin, I., & Sari, Y. A. (2021). Pengaruh Metode Word Embedding dalam Vector Space Model pada Pemerolehan Informasi Materi IPA Siswa SMP (Vol. 5, Issue 3). <http://j-ptiik.ub.ac.id>