

**DETEKSI UJARAN KEBENCIAN DALAM DOMAIN POLITIK
PADA MEDIA SOSIAL DENGAN ALGORITMA
LONG SHORT-TERM MEMORY (LSTM)**

TESIS

Oleh:

**ALFIANDRI PUTRA PERDANA
204211047**



**PROGRAM STUDI S-2 TEKNOLOGI INFORMASI
FAKULTAS INFORMATIKA
UNIVERSITAS MIKROSKIL
MEDAN
2023**

**HATE SPEECH DETECTION IN POLITICAL DOMAINS ON
SOCIAL MEDIA USING LONG SHORT-TERM MEMORY (LSTM)**

THESIS

By:

**ALFIANDRI PUTRA PERDANA
204211047**



**UNIVERSITAS
MIKROSKIL**

**MAJOR OF S-2 INFORMATION TECHNOLOGY
FACULTY OF INFORMATICS
UNIVERSITAS MIKROSKIL
MEDAN
2023**

LEMBARAN PENGESAHAN

DETEKSI UJARAN KEBENCIAN DALAM DOMAIN POLITIK
PADA MEDIA SOSIAL DENGAN ALGORITMA
LONG SHORT-TERM MEMORY (LSTM)

TESIS

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Magister
Program Studi S-2 Teknologi Informasi

Oleh:

ALFIANDRI PUTRA PERDANA
204211047

Disetujui Oleh:


Dosen Pembimbing,



Dr. Ronsen Purba, M.Sc.

Medan, 27 Januari 2023
Diketahui dan Disahkan Oleh:

Ketua Program Studi
S-2 Teknologi Informasi



Ir. Erwin Setiawan Panjaitan, M.M.S.I., Ph.D.

HALAMAN PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Program Studi S-2 Teknologi Informasi Universitas Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Alfiandri Putra Perdana
NIM : 204211047
Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul “DETEKSI UJARAN KEBENCIAN DALAM DOMAIN POLITIK PADA MEDIA SOSIAL DENGAN ALGORITMA *LONG SHORT-TERM MEMORY* (LSTM)”, dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak menyuruh orang lain yang mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar. Bila di kemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya (membuatnya), maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh Universitas Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada Universitas Mikroskil Medan Hak Bebas Royalti Non-eksklusif (Non-exclusive Royalty Free Right) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, Universitas Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tesis saya, secara keseluruhan atau hanya sebagian atau hanya ringkasannya saja dalam bentuk format tercetak dan/atau elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak/sistem informasi.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 27 Januari 2023

Saya yang membuat pernyataan,



Alfiandri Putra Perdana

DETEKSI UJARAN KEBENCIAN DALAM DOMAIN POLITIK PADA MEDIA SOSIAL DENGAN ALGORITMA LONG SHORT-TERM MEMORY (LSTM)

Abstrak

Maraknya ujaran kebencian dalam domain politik di media sosial dapat memicu tindakan kekerasan, prasangka negatif bahkan perpecahan. Dengan metode pendeteksian ujaran kebencian dapat membantu menganalisis ujaran antara ujaran kebencian dan bukan ujaran kebencian. Namun, akurasi deteksi ujaran kebencian dibatasi oleh kualitas sumber dataset, penggunaan bahasa informal, definisi yang berbeda tentang apa yang dimaksud dengan ujaran kebencian. Penelitian ini memberikan solusi atas permasalahan ini menggunakan metode Long Short-Term Memory dan teknik pembobotan kata TF-IDF. Dataset yang digunakan akan melalui beberapa tahapan pre-processing, seperti Data Cleaning, Case Folding, Tokenizing, Filtering, dan Stemming. Setelah itu dataset akan dilakukan pembobotan dengan teknik TF-IDF. Kemudian seluruh data dilatih dan divalidasi dengan menggunakan model LSTM. Penelitian ini menghasilkan 12 model yang dilatih menggunakan dataset Alfina, Ibrohim, dataset hasil scrapping penulis, serta gabungan dataset tersebut yang kemudian dapat digunakan untuk menganalisis ujaran kebencian. Hasil pengujian menunjukkan proses deteksi ujaran kebencian lebih akurat, peningkatan akurasi dari 74,47% menjadi 96,58% dibandingkan dengan pengujian tanpa menerapkan metode usulan.

Kata kunci: Deteksi Ujaran Kebencian Dalam Domain Politik, Long Short-Term Memory, Term Frequency - Inverse Document Frequency

Abstract

The spread of hate speech in the political domain on social media can trigger acts of violence, negative prejudice and even division. With the hate speech detection method, it can help analyze speech between hate speech and non-hate speech. However, the accuracy of hate speech detection is limited by the quality of the dataset sources, use of informal language, different definitions of what is meant by hate speech. This research provides a solution to this problem using the Long Short-Term Memory method and the TF-IDF word weighting technique. The dataset used will go through several pre-processing stages, such as Data Cleaning, Case Folding, Tokenizing, Filtering, and Stemming. After that the dataset will be weighted using the TF-IDF technique. Then all data is trained and validated using the LSTM model. This study produced 16 models that were trained using the Alfina, Ibrohim dataset, the author's scrapping dataset, and a combination of these datasets which can then be used to analyze hate speech. The test results show that the hate speech detection process is more accurate, increasing accuracy from 74.47% to 96.58% compared to the test without applying the proposed method.

Keywords: Hate Speech Detection in Political Domain, Long Short-Term Memory, Term Frequency - Inverse Document Frequency

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “DETEKSI UJARAN KEBENCIAN DALAM DOMAIN POLITIK PADA MEDIA SOSIAL DENGAN ALGORITMA LONG SHORT-TERM MEMORY”.

Tesis ini disusun untuk melengkapi persyaratan kurikulum guna menyelesaikan Program Studi Magister Teknologi Informasi Strata Dua Universitas Mikroskil Medan.

Tesis ini tidak akan berjalan dengan baik tanpa adanya bimbingan, nasehat, dan dukungan dari berbagai pihak yang telah banyak meluangkan waktu, tenaga, dan pikiran. Oleh karena itu, penulis menyampaikan banyak terima kasih kepada.

1. Bapak Dr. Ronsen Purba, M.Sc., selaku Dosen Pembimbing I yang telah membimbing penulis dalam proses penyelesaian Tesis ini.
2. Bapak Muhammad Fermi Pasha, B.Sc., M.Sc., Ph.D., selaku Pendamping Pembimbing yang telah membimbing penulis dalam proses penyelesaian Tesis ini.
3. Bapak Hardy, S.Kom., M.Sc., Ph.D., selaku Rektor Universitas Mikroskil Medan.
4. Bapak Ng Poi Wong, S.Kom., M.T.I., selaku Dekan Fakultas Informatika Universitas Mikroskil Medan.
5. Bapak Ir. Erwin Setiawan Panjaitan M.M.S.I., Ph.D., selaku Ketua Program Studi S-2 Teknologi Informasi Fakultas Informatika Universitas Mikroskil Medan.
6. Bapak/Ibu Dosen Penguji yang sudah memberikan pelajaran kepada penulis.
7. Orang tua khususnya Mama yang selalu memberikan dukungan kepada penulis.
8. Istri tercinta, Siska Elvira Sari Hasibuan, S.Kom., yang selalu memberikan dukungan secara moril dan selalu memotivasi penulis.
9. Ananda Arshya Devanka Alfiska yang menjadi salah satu motivasi penulis.
10. Semua pihak yang telah membantu penulis dalam penulisan Tesis ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan Tesis ini masih terdapat banyak kekurangan yang ada. Oleh sebab itu, penulis sangat mengharapkan dan menghargai kritik dan saran yang bersifat membangun akan sangat diterima. Akhir kata, penulis berharap semoga penyusunan Tesis ini dapat bermanfaat bagi masyarakat. Terima kasih.

Medan, 27 Januari 2023

Penulis,



Alfiandri Putra Perdana



UNIVERSITAS MIKROSKIL

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	ix
DAFTAR LAMPIRAN	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat	3
1.5 Ruang Lingkup	3
1.6 Metodologi Penelitian	4
Bab II KAJIAN LITERATUR	5
2.1 Tinjauan Pustaka	5
2.1.1 Ujaran Kebencian	5
2.1.2 Ujaran Kebencian di <i>Twitter</i>	8
2.1.3 Pre-processing	9
2.1.4 Recurrent Neural Network	11
2.1.5 Long Short-Term Memory	12
2.1.6 TF-IDF	14
2.1.7 Confusion Matrix	14
2.2 Penelitian Sebelumnya	16
2.3 Kerangka Konsep	18
Bab III METODOLOGI PENELITIAN	19
3.1 Analisis Masalah	19
3.2 Data Yang Dibutuhkan	19
3.3 Metode Penelitian	21
3.4 Alat-Alat Penelitian	29
3.5 Teknik Analisis	29
3.6 Jadwal Penelitian	30
Bab IV HASIL DAN PEMBAHASAN	31

4.1	Hasil	31
4.1.1	Pengambilan Data	31
4.1.2	<i>Pre-processing</i>	31
4.1.3	Pelatihan dan Validasi Model	37
4.1.4	Deteksi Ujaran Kebencian	48
4.1.5	Perbandingan Dengan Metode Lain	49
4.2	Pembahasan	49
4.2.1	Pengaruh Dataset terhadap Model	49
4.2.2	Pengaruh Pra-pemrosesan terhadap Hasil Deteksi	50
4.2.3	Pengaruh Teknik Pembobotan TF-IDF terhadap Hasil Deteksi	50
4.2.4	Perbandingan Hasil Dengan Metode Lain	50
Bab V	PENUTUP	51
5.1	Kesimpulan	51
5.2	Saran	51
DAFTAR PUSTAKA		52
DAFTAR RIWAYAT HIDUP		55



UNIVERSITAS MIKROSKIL

DAFTAR GAMBAR

Gambar 2.1 Contoh Ujaran Kebencian di Twitter.....	9
Gambar 2.2 Pra-pemrosesan Data	11
Gambar 2.3 Struktur tradisional Recurrent Neural Networks (RNN).....	12
Gambar 2.4. Arsitektur Long Short-Term Memory	13
Gambar 2.5. Kerangka Konsep Penelitian.....	18
Gambar 3.1 Flowchart Metode Penelitian.....	23
Gambar 3.2 Flowchart Tahapan Pra-pemrosesan.....	25
Gambar 3.3 Flowchart Proses Pembuatan dan Pelatihan Model.....	26
Gambar 3.4 Flowchart Pelatihan LSTM.....	28
Gambar 3.5 Flowchart Tahapan Deteksi	29
Gambar 3.6 Contoh Confusion Matrix	30
Gambar 4.1 Hasil Validasi Model Pengujian Alfina Tanpa TF-IDF dan Pra-pemrosesan.....	37
Gambar 4.2 Confusion Matrix Pengujian Alfina Tanpa TF-IDF dan Pra-pemrosesan.....	38
Gambar 4.3 Hasil Validasi Model Pengujian Alfina Tanpa TF-IDF Dengan Pra-pemrosesan.....	38
Gambar 4.4 Confusion Matrix Pengujian Alfina Tanpa TF-IDF Dengan Pra-pemrosesan.....	38
Gambar 4.5 Hasil Validasi Model Pengujian Alfina Dengan TF-IDF dan Pra-pemrosesan.....	39
Gambar 4.6 Confusion Matrix Pengujian Alfina Dengan TF-IDF dan Pra-pemrosesan.....	39

Gambar 4.7 Hasil Validasi Model Pengujian Ibrohim Tanpa TF-IDF dan Pra-pemrosesan.....	40
Gambar 4.8 Confusion Matrix Pengujian Ibrohim Tanpa TF-IDF dan Pra-pemrosesan.....	40
Gambar 4.9 Hasil Validasi Model Pengujian Ibrohim Tanpa TF-IDF dan Dengan Pra-pemrosesan	40
Gambar 4.10 Confusion Matrix Pengujian Ibrohim Tanpa TF-IDF dan Dengan Pra-pemrosesan	41
Gambar 4.11 Hasil Validasi Model Pengujian Ibrohim Dengan TF-IDF dan Pra-pemrosesan.....	41
Gambar 4.12 Confusion Matrix Pengujian Ibrohim Dengan TF-IDF dan Pra-pemrosesan.....	42
Gambar 4.13 Hasil Validasi Model Pengujian Data Realtime Tanpa TF-IDF dan Pra-pemrosesan	42
Gambar 4.14 Confusion Matrix Pengujian Data Realtime Tanpa TF-IDF dan Pra-pemrosesan.....	42
Gambar 4.15 Hasil Validasi Model Pengujian Data Realtime Tanpa TF-IDF dan Dengan Pra-pemrosesan.....	43
Gambar 4.16 Confusion Matrix Pengujian Data Realtime Tanpa TF-IDF dan Dengan Pra-pemrosesan	43
Gambar 4.17 Hasil Validasi Model Pengujian Data Realtime Dengan TF-IDF dan Pra-pemrosesan	43
Gambar 4.18 Confusion Matrix Pengujian Data Realtime Dengan TF-IDF dan Pra-pemrosesan.....	44

Gambar 4.19 Hasil Validasi Model Pengujian Gabungan Tanpa TF-IDF dan Pra-pemrosesan.....	45
Gambar 4.21 Confusion Matrix Pengujian Gabungan Tanpa TF-IDF dan Pra-pemrosesan.....	45
Gambar 4.22 Hasil Validasi Model Pengujian Gabungan Tanpa TF-IDF dan Dengan Pra-pemrosesan	45
Gambar 4.23 Confusion Matrix Pengujian Gabungan Tanpa TF-IDF dan Dengan Pra-pemrosesan	46
Gambar 4.24 Hasil Validasi Model Pengujian Data Realtime Dengan TF-IDF dan Pra-pemrosesan	46
Gambar 4.25 Confusion Matrix Pengujian Gabungan Dengan TF-IDF dan Pra-pemrosesan.....	47



UNIVERSITAS MIKROSKIL

DAFTAR TABEL

Tabel 2.1 Contoh Ujaran Kebencian pada Domain Politik	7
Tabel 2.2 Pra-pemrosesan Data Tweet	10
Tabel 2.3 Confusion Matrix.....	15
Tabel 2.4 Penelitian Sebelumnya Terkait Pendeteksian Ujaran Kebencian	16
Tabel 3.1 Struktur Dataset Penelitian Alfina, I., Fanany, M. I., dan Mulia, R.....	19
Tabel 3.2 Struktur Dataset Penelitian Ibrohim, M. O. dan Budi, I.....	20
Tabel 3.3 Struktur Dataset Real-time Twitter	21
Tabel 3.4 Jadwal Penelitian	30
Tabel 4.1 Konversi Data Label Menjadi Bentuk Numerik.....	31
Tabel 4.2 Hasil Data Cleaning.....	32
Tabel 4.3 Hasil Case Folding	33
Tabel 4.4 Perubahan Kata Informal Menjadi Formal.....	34
Tabel 4.5 Hasil Tokenizing	34
Tabel 4.6 Hasil Filtering Stopwords.....	35
Tabel 4.7 Hasil Stemming	36
Tabel 4.8 Hasil Pelatihan dan Akurasi Deteksi	47
Tabel 4.9 Contoh Ujaran Yang Akan Dideteksi.....	48
Tabel 4.10 Perbandingan Dengan Metode Lain	49

DAFTAR LAMPIRAN

Lampiran 1: DAFTAR RIWAYAT HIDUP	5
----------------------------------------	---



UNIVERSITAS MIKROSKIL