

**Identifikasi komentar ujaran kebencian pada media sosial indonesia
menggunakan metode kombinasi RNN-LSTM untuk identifikasi komentar
ujaran kebencian pada data tweet**

TUGAS AKHIR

Diajukan Untuk Memenuhi Persyaratan
Guna Meraih Gelar Sarjana
Informatika Universitas Muhammadiyah Malang



**NAMA: SUKMA
NIM: 201910370311337**

**Bidang Minat
(Data Science)**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG**

2024

LEMBAR PERSETUJUAN

**Identifikasi komentar ujaran kebencian pada media sosial indonesia
menggunakan metode kombinasi RNN dan LSTM untuk identifikasi
komentar ujaran kebencian pada data tweet**

TUGAS AKHIR

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Stara 1 Teknik
Informatika Universitas Muhammadiyah Malang**

Menyetujui,
Malang, 7 Juli 2024

Pembimbing I



Nur Hayatin, S.Kom, M.Kom
NIDN 0726038402

Pembimbing II



Yufis Azhar, S.Kom, M.Kom
NIP.108.1410.0544

LEMBAR PENGESAHAN

Identifikasi komentar ujaran kebencian pada media sosial indonesia menggunakan metode kombinasi RNN dan LSTM untuk identifikasi komentar ujaran kebencian pada data tweet

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

SUKMA

201910370311337

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian pada tanggal 7 Juli 2024

Menyetujui,

Dosen Penguji 1



Christian Sri Kusuma Aditya S.Kom.,

M.Kom

NIP. 180327021991PNS.

Dosen Penguji 2



Setio Basuki MT., Ph.D.

NIP. 10809070477PNS.

Mengetahui,
Ketua Jurusan Informatika



Ir. Galih Wasfa Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : SUKMA

NIM : 201910370311337

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“Identifikasi komentar ujaran kebencian pada media sosial indonesia menggunakan metode kombinasi RNN dan LSTM untuk identifikasi komentar ujaran kebencian pada data tweet”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Nur Hayatin S. ST., M.Kom

Malang, 7 Juli 2024
Yang Membuat Pernyataan



SUKMA

ABSTRAK

Ujaran kebencian adalah perlakuan kebencian atau perundungan secara daring, Tindakan penyerangan, penghinaan, atau menyakiti orang lain secara berulang-ulang yang terjadi di media sosial, contoh kasus yaitu Twitter. Twitter merupakan salah satu media pertukaran informasi yang mudah penggunaannya dan sangat populer, twitter pun salah satu media sosial yang banyak digunakan Indonesia karena memiliki persebaran informasi yang sangat cepat. . Penelitian ini mengusulkan mengidentifikasi ujaran kebencian dari media sosial berbahasa Indonesia dengan menggunakan pendekatan hybrid deep learning. Kombinasi recurrent neural network(RNN) dan long short term memor. RNN dapat memetakan urutan vector masukan berukuran tetap ke komponen vector tersembunyi LSTM diimplementasikan dengan mempertimbangkan komponen pertumbuhan vector gradien yang mungkin ada di RNN Langkah-langkah yang dilakukan meliputi preprocessing, pemodelan, implementasi, dan evaluasi. Dataset yang digunakan berasal dari Bahasa Indonesia yang berjumlah 13158, dataset tersebut merupakan dari Kaggle, dataset di bagi menjadi dua jelas yaity tweet dan ujaran kebencian. Hasil dari evaluasi menunjukkan model yang di usulkan menghasilkan f-measure sebesar 86%.

Kata kunci : klasifikasi, hybrid deep learning, ujaran kebencian.

ABSTRACT

Hate speech is hate speech or online bullying, repeated acts of attacking, insulting or hurting other people that occur on social media, an example of which is Twitter. Twitter is a medium for exchanging information that is easy to use and very popular. Twitter is also one of the social media that is widely used in Indonesia because it spreads information very quickly. . This research proposes identifying hate speech from Indonesian language social media using a hybrid deep learning approach. Combination of recurrent neural network(RNN) and long short term memory. RNN can map fixed-sized input sequences to hidden vector components. LSTM is implemented by considering gradient vector growth components that may exist in the RNN. The steps taken include preprocessing, modeling, implementation, and evaluation. The dataset used comes from. There are 13,158 Indonesian languages, the dataset is from Kaggle, the dataset is divided into two categories, namely tweets and hate speech. The results of the evaluation show that the proposed model produces an f-measure of 86%.

Keywords: classification, hybrid deep learning, hate speech.

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul :
”Identifikasi Komentar Ujaran Kebencian Pada Media Sosial Menggunakan Metode Kombinasi RNN-LSTM Untuk Identifikasi Ujaran Kebencian pada Data Tweet”

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi pedahuluan, kajian pustaka, metode penelitian, hasil dan pembahasan yang telah penulis dapatkan dari hasil penelitian dan simpulan yang telah di dapatkan oleh penulis. Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan arahan yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 22 Juni 2024



DAFTAR ISI

LEMBAR PENGESAHAN	3
LEMBAR PERNYATAAN	4
ABSTRAK.....	5
ABSTRACT.....	6
LEMBAR PERSEMBAHAN	7
KATA PENGANTAR	10
BAB I.....	16
PENDAHULUAN	16
1.2 Rumusan Masalah.....	18
1.3 Tujuan Penelitian.....	18
1.4 Batasan Masalah	18
BAB II TINJAUAN PUSTAKA	19
2.1 Penelitian Terdahulu	19
2.4 Long Short-Term Memory(LSTM).....	21
2.5 Word Embedding	23
2.6 Natural Language Toolkit(NLTK).....	24
2.7 Rectified Linear Units(ReLU).....	24
2.8 Layer Dense.....	25
2.9 Layer Dropout.....	25
2.10 Optimasi Adam.....	27
BAB III.....	28
3.1 Dataset	28
3.2 Preprocessing	29
3.3 Data Splitting	29
3.4 Kombinasi RNN-LSTM.....	29
3.6 Implementasi Model Menggunakan RNN-LSTM	31
BAB IV.....	34
4.1 Data Cleaning & Pre-processing	34
4.1.1 Case Folding	34
4.1.2 Stopword Removal.....	34
4.1.3 Normalisasi	35
4.1.4 Stemming	35
4.2 Proses Klasifikasi	36
4.2.1 Tokenisasi	36
4.2.2 Sequencing and Padding	37

4.3	Build Model	37
4.4	Evaluasi Model.....	38
4.4.1	Model Accuracy Plot.....	38
4.4.2	Model Loss Plot	39
4.4.3	Confussion Matrix	40
BAB V KESIMPULAN.....		48
Daftar Pustaka.....		49



DAFTAR GAMBAR

Gambar 2. 1 Susunan Layer(a) sebelum penggunaan dropout (b) Sesudah penggunaan dropout.....	26
Gambar 2. 2 (a) Unit pada waktu pelatihan (b) Unit pada saat pengujian	26
Gambar 3. 1 Metodologi Indetifikasi Ujaran Kebencian Pada Media Sosial	28
Gambar 3. 2 Konsep RNN.....	30
Gambar 3. 3 Arsitektur LSTM.....	30
Gambar 3. 4 Implementasi RNN-LSTM dalam sebuah model.....	32
Gambar 4. 1 Source Code Stopword Removal	34
Gambar 4. 2 Source Code Normalisasi.....	35
Gambar 4. 3 Source Code Stemming.....	36
Gambar 4. 4 Source Code Tokenisasi	36
Gambar 4. 5 Source Code Sequencing.....	37
Gambar 4. 6 Source Code Padding	37
Gambar 4. 7 Source Code Build Model.....	38
Gambar 4. 8Source Code Plotting Model Akurasi	38
Gambar 4. 9 Hasil Plotting Model Akurasi.....	39
Gambar 4. 10 Source Code Plotting Model Loss.....	39
Gambar 4. 11 Hasil Plotting Model Loss.....	40
Gambar 4. 12 Confussion Matrix	41
Gambar 4. 13 Source Code Kemungkinan sistem salah prediksi.....	42
Gambar 4. 14 Source Code Uji Hate Speech dan Non- Hate Speech Pada Tweet.....	42

DAFTAR TABLE

Table 2. 1 Penelitian Terdahulu.....	20
Table 4. 1 Hasil Uji Metode	42



DAFTAR GRAFIK

Grafik 4. 1 Hasil Plotting Model Akurasi	39
Grafik 4. 2 Hasil Plotting Model Loss	40



no. 2, 2020.

[11] S. Suherwinto, "ANALISIS GEOSTRATEGIS INDONESIA SEBAGAI SENTRAL GERAKAN TERORISME DI ASIA TENGGARA: DARI JAMAAH ISLAMİYAH SAMPAI JAMAAH ANSHARUT DAULAH,"

Jakarta, 2019. [Daring]. Tersedia pada: <https://indonesiadefenseanalysis.blogspot.com/2011/02/bali-bombing-series.html>.

[12] F. A. PRADANA, "skripsi," PERBANDINGAN WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT MENGGUNAKAN DEEP LEARNING PADA ULASAN KONDISI PENGGUNA OBAT KESEHATAN , 2023.

[13] F. F. Ahmad O, "Jurnal Skripsi," ANALISIS UJARAN KEBENCIAN DALAM MEDIA SOSIAL TWITTER , 2005.

[14] K. M. Fitria, "Skripsi," IMPLEMENTASI MODEL BAHASA OPENAI GPT-3 UNTUK APLIKASI TEXT CONTENT GENERATOR BERBASIS WEB DAN APLIKASI MOBILE, 2022.

[15] M. Furqan, "Skripsi," Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia Sentiment Analysis using K-Nearest Neighbor towards the New Normal During the Covid-19 Period in Indonesia, 2022.

[16] D. T. H. M. R. Adherda, "Jurnal," GENDER CLASSIFICATION BASED ON VOICE USING RECURRENT NEURAL NETWORK (RNN), 2023.

[17] I. Z. M. Putra, "Jurnal," Model Klasifikasi Berbasis Multiclass Classification dengan Kombinasi Indobert Embedding dan Long Short-Term Memory untuk Tweet Berbahasa Indonesia (Classification Model Based on Multiclass Classification with a Combination of Indobert Embedding and Long , 2022.

[18] B. Putrama, "Jurnal," Analisis Kinerja Long Short-Term Memory, Peephole Connection LSTM dan Facebook's Prophet dalam Memprediksi Pergerakan Harga Saham PT Telekomunikasi Indonesia Tbk, 2021.

[19] P. H. ARDI, "Skripsi," PERAMALAN PERMINTAAN EKSPOR NONMIGAS INDONESIA MENGGUNAKAN LONG

SHORT TERM MEMORY, 2023.

[20] S. R. Silvanie Astried, "jurnal," APLIKASI CHATBOT UNTUK FAQ AKADEMIK DI IBI-K57 DENGAN LSTM DAN PENYEMATAN KATA, 2022.

[21] R. Edoardo, "jurnal," Text Summarization Menggunakan Library Natural Language Toolkit (NLTK) Berbasis Pemrograman Python, 2020.

[22] P. C. S. GULO, "Skripsi," ANALISIS SENTIMEN KULIAH ONLINE SELAMA PANDEMI COVID-19 MENGGUNAKAN ALGORITMA NAIVE BAYES, 2021.

[23] F. F. TALININGSIH, "Jurnal," Sistem Otentikasi Biometrik Berbasis Sinyal EKG Menggunakan Convolutional Neural Network 1 Dimensi, 2022

[24] S. Kemp, "Digital in Indonesia: All the Statistics You Need in 2021 — DataReportal — Global Digital Insights," 2021. <https://datareportal.com/reports/digital-2021-indonesia> (accessed Jun. 27, 2021).

[25] A. Briliani, B. Irawan, and C. Setianingsih, "Hate speech detection in Indonesian language on Instagram comment section using K-nearest neighbor classification method," Proc. - 2019 IEEE Int. Conf. Internet Things Intell. Syst. IoTaIS 2019.

[26] S. Y. Hukmana, "125 Akun Medsos Terjaring Virtual Police - Medcom.id," 2020. <https://www.medcom.id/nasional/hukum/gNQ5RnwN-125-akun-medsos-terjaring-virtual-police> (accessed Jun. 27, 2021).

[27] L. P. A. S. Tjahyanti, "Pendeteksian Bahasa Kasar (Abusive Language) Dan Ujaran Kebencian (Hate Speech) Dari Komentar Di Jejaring Sosial," J. Chem. Inf. Model., vol. 07, no. 9, pp. 1689–1699, 2020.

[28] P. Devita, "Apakah semua ujaran kebencian perlu dipidana? Catatan untuk revisi UU ITE." (accessed Jun. 27, 2021).

[29] F. Alzami, N. P. P, R. A. P, R. A. Megantara, and D. P. Prabowo, "SENTIMENT ANALYSIS UNTUK DETEKSI UJARAN KEBENCIAN PADA DOMAIN POLITIK," vol. 5, no. Sens 5, pp. 213–218, 2020.

[30] K. M. Hana, Adiwijaya, S. Al Faraby, and A. Bramantoro, "Multi-label Classification of Indonesian Hate Speech on Twitter Using Support Vector Machines," 2020 Int. Conf. Data Sci. Its Appl. ICoDSA 2020, 2020, doi:

[31] G. T. Indrawan, "Jurnal," Diagnosis COVID-19 Berdasarkan Citra X-ray Paru-Paru Menggunakan Metode Convolutional Neural Network , 2022.

[32] H. C. HARISA, "Skripsi," PENERAPAN ALGORITMA DENSE NEURAL NETWORK UNTUK KLASIFIKASI JENIS SEDIMEN DASAR LAUT MENGGUNAKAN DATA MOSAIK DAN POLA FITUR BACKSCATTER , 2023.

[33] A. Rabiatul, "Jurnal," EKSPLORASI KAPASITAS PENGKODEAN AMPLITUDO UNTUK MODEL QUANTUM MACHINE LEARNING, 2023.

[34] R. P. Iham, "Jurnal," JURNAL MEDIA INFORMATIKA BUDIDARMA Pengenalan Potensi Racun dan Peningkatan Keamanan Pangan Dalam Jamur Menggunakan Convolutional Neural Network, 2023.

[35] S. Nitish, "Jurnal," Dropout: A Simple Way to Prevent Neural Networks from Overfitting, 2014.

[36] I. Rafiedhia Pramutighna, "Jurnal," Pengenalan Potensi Racun dan Peningkatan Keamanan Pangan Dalam Jamur Menggunakan Convolutional Neural Network, 2023.

[37] H. Darmanto, "Jurnal," PENGENALAN SPESIES IKAN BERDASARKAN KONTUR OTOLITH MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK, 2019.



UNIVERSITAS
MUHAMMADIYAH
MALANG



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : SUKMA
NIM : 201910370311337
Judul TA : Identifikasi komentar ujaran kebencian pada media sosial indonesia menggunakan metode kombinasi RNN-LSTM untuk identifikasi komentar ujaran kebencian pada data tweet

Hasil Cek Plagiarisme dengan Turnitin

No	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	8%
2.	Bab 2 – Daftar Pustaka	25 %	3%
3.	Bab 3 – Analisis dan Perancangan	25 %	6%
4.	Bab 4 – Implementasi dan Pengujian	15 %	4%
5.	Bab 5 – Kesimpulan dan Saran	5 %	4%
6.	Makalah Tugas Akhir	20%	9%

*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

