

SKRIPSI

DETEKSI BERITA HOAX MENGGUNAKAN LSTM (LONG SHORT TERM MEMORY) DENGAN BASIS REPRESENTASI TF-IDF (TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY)



Oleh :

STEVEN ALEXANDER

190411100131

Dosen Pembimbing 1: Mula'ab, S.Si., M.Kom

Dosen Pembimbing 2: Fifin Ayu Mufarroha, S.Kom., M.Kom

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA**

2024

HALAMAN JUDUL

**DETEKSI BERITA HOAX MENGGUNAKAN LSTM (LONG SHORT
TERM MEMORY) DENGAN BASIS REPRESENTASI
TF-IDF (TERM FREQUENCY – INVERSE
DOCUMENT FREQUENCY)**

SKRIPSI

**Diajukan untuk Memenuhi Persyaratan Penyelesaian
Studi Strata Satu (S1) dan Memperoleh Gelar Sarjana Komputer (S.Kom) di
Universitas Trunojoyo Madura**

Oleh:
Steven Alexander
190411100081

**PROGRAM STUDI S1 TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA
2024**

HALAMAN PENGESAHAN

DETEKSI BERITA HOAX MENGGUNAKAN LSTM (LONG SHORT
TERM MEMORY) DENGAN BASIS REPRESENTASI
TF-IDF (TERM FREQUENCY - INVERSE
DOCUMENT FREQUENCY)

Skripsi diajukan sebagai salah satu syarat untuk menyelesaikan pendidikan
program studi S1 Teknik Informatika Universitas Trunojoyo Madura

Oleh:

Nama : Steven Alexander

NIM : 190411100131

Disetujui oleh Tim Penguji Skripsi: Tanggal Sidang 05 Juni 2024

Mula'ab, S.Si., M.Kom.

(Pembimbing I)

NIP. 19730520 200212 1 001

Fifin Ayu Mufarroha, S. Kom., M.Kom.

(Pembimbing II)

NIP. 19891011 202012 2 010

Andharini Dwi Cahyani, S. Kom., M.Kom., Ph.D

(Penguji I)

NIP. 19800503 200312 2 002

Dr. Eka Mala Sari Rochman, S.Kom.M.Kom.

(Penguji II)

NIP. 19840716 200812 2 001

Yoga Dwitva Pramudita, S.Kom., M.Cs.

(Penguji III)

NIP. 19840413 200812 1 002

Bangkalan, 29 Oktober 2024

Mengetahui

Ketua Jurusan Teknik Informatika

Dr. Yeni Kustivahningsih, S.Kom., M.Kom.

NIP. 19770921 200812 2 002

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan dibawah ini:

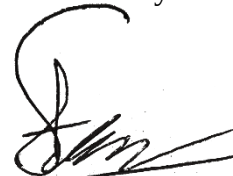
Nama : Steven Alexander
NIM : 190411100131
Judul Tugas Akhir : Deteksi Berita Hoax Menggunakan
LSTM (Long Short Term Memory)
dengan Basis Representasi TF-IDF
(Term Frequency Inverse Document
Frequency)

Menyatakan dengan sebenarnya bahwa penulisan tugas akhir ini berdasarkan hasil penelitian, pemikiran dan pemaparan saya sendiri, bukan merupakan karya pihak lain serta belum pernah diajukan untuk mendapatkan gelar akademik Sarjana Komputer baik dalam lingkungan Universitas Trunojoyo Madura maupun di Perguruan Tinggi lain. Jika terdapat karya orang lain sebagai referensi pada penelitian ini, maka saya akan mencantumkan sumber yang jelas.

Dengan pernyataan ini saya buat dengan sesungguhnya dan apabila terbukti tugas akhir ini sebagian atau seluruhnya merupakan hasil plagiasi atau terdapat hal-hal yang tidak sesuai dengan pernyataan diatas, saya sanggup menerima sanksi akademis sesuai peraturan yang berlaku di Universitas Trunojoyo Madura.

Bangkalan, 29 Oktober 2024

Pembuat Pernyataan,



Steven Alexander
NIM.190411100131

ABSTRAK

Penyebaran berita *hoax* telah menjadi ancaman serius dalam era informasi digital saat ini. Penyebaran berita *hoax* atau berita palsu dapat memiliki dampak yang merugikan dan berbahaya, baik bagi individu maupun masyarakat secara keseluruhan. Penyebaran Berita palsu dapat mempengaruhi persepsi dan pemahaman orang terhadap suatu peristiwa atau isu. Tujuan penelitian ini adalah untuk menerapkan sebuah model yang efektif dalam mengenali dan membedakan teks berita *hoax* dari teks berita *non-hoax*. metode TF-IDF digunakan untuk menghasilkan *Vector Space Model* dari teks berita, dengan mempertimbangkan frekuensi kemunculan kata dan bobotnya dalam korpus keseluruhan. Kemudian, arsitektur LSTM dipakai sebagai model klasifikasi yang mampu memahami konteks jangka panjang dalam teks. Tahapan penelitian meliputi pengumpulan data berita *hoax* dan *non-hoax* dari Kaggle (*fake.csv* dan *true.csv*), preprocessing teks (membersihkan dan mengubah teks menjadi representasi vektor), pelatihan model LSTM, dan evaluasi kinerja model menggunakan confusion matrix. Setelah pengujian, kombinasi parameter LSTM dengan akurasi tertinggi sebesar 99,65% dicapai pada skenario uji coba 14 (80:20, neuron hidden layer 16, epoch 100, kompilasi model dengan *binary crossentropy* sebagai fungsi *loss*, adam sebagai *optimizer*, dan metrik akurasi untuk evaluasi) dengan penerapan preprocessing yakni *tokenizing*, *stopword removal*, dan *stemming*. Hasil ini menunjukkan bahwa model klasifikasi yang dikembangkan sangat akurat dalam mendeteksi berita *hoax*

Kata kunci: klasifikasi teks, berita *hoax*, TF-IDF, LSTM, deteksi berita *hoax*

ABSTRACT

The spread of hoax news has become a serious threat in today's digital information era. The dissemination of hoax news or false news can have harmful and dangerous impacts, both for individuals and society as a whole. The spread of fake news can influence people's perceptions and understanding of an event or issue. The aim of this research is to implement an effective model to recognize and distinguish hoax news texts from non-hoax news texts. The TF-IDF method is used to generate a Vector Space Model from the news texts, considering the frequency of word occurrences and their weights in the overall corpus. Then, the LSTM architecture is employed as a classification model capable of understanding long-term context in the text. The research stages include collecting hoax and non-hoax news data from Kaggle (fake.csv and true.csv), preprocessing the text (cleaning and converting the text into vector representation), training the LSTM model, and evaluating the model's performance using a confusion matrix. After testing, the combination of LSTM parameters with the highest accuracy of 99.65% was achieved in trial scenario 14 (80:20 split, hidden layer neurons 16, 100 epochs, compiling the model with binary crossentropy as the loss function, Adam as the optimizer, and accuracy as the evaluation metric) with preprocessing applied, including tokenizing, stopwords removal, and stemming. These results show that the developed classification model is highly accurate in detecting hoax news.

Key word: text classification, hoax news, TF-IDF, LSTM, hoax news detection

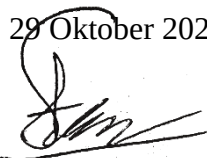
KATA PENGANTAR

Pujan dan Ucapan Syukur penulis panjatkan kepada Tuhan Yesus Kristus atas anugerah dan penyertaan-Nya, penulis dapat menyelesaikan skripsi yang berjudul **“DETEKSI BERITA HOAX MENGGUNAKAN LSTM (LONG SHORT TERM MEMORY) DENGAN BASIS REPRESENTASI TF-IDF (TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY)”**. Penulis juga tidak lupa mengucapkan terimakasih sebesar-besarnya atas terselesainya skripsi ini kepada:

1. Roh Kudus yang telah menolong dan menuntun umat manusia dalam mengikuti ajaran kasih
2. Orang Tua serta keluarga besar yang telah mendoakan, memberikan motivasi, dan semangat kepada penulis.
3. Bapak Mula'ab, S.Si., M.Kom. selaku dosen pembimbing I dan Ibu Fifin Ayu Mufarroha, S. Kom., M.Kom. selaku dosen pembimbing II. Terima kasih atas waktu dan tenaga yang diluangkan kepada penulis selama proses pengerjaan skripsi ini.
4. Bapak dan Ibu Penguji yang telah membimbing dalam proses penyelesaian skripsi ini.
5. Bapak dan Ibu Dosen Fakultas Teknik, khususnya Prodi Informatika yang telah mengajarkan materi-materi tentang informatika selama masa perkuliahan.
6. Serta teman-teman seperjuangan, terima kasih atas segala hal yang telah diberikan selama menjalani perantauan sebagai mahasiswa di Universitas Trunojoyo Madura. Semoga kalian semua menjadi orang-orang sukses dan berguna untuk memajukan bangsa.

Penulis menyadari bahwa masih banyak kekurangan dalam penulisan skripsi ini. Oleh karena itu, penulis memohon maaf dan menerima segala kritik dan saran yang membangun dari rekan-rekan semua.

Bangkalan, 29 Oktober 2024



Steven Alexander

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
DAFTAR KODE PROGRAM.....	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.2.1 Permasalahan.....	3
1.2.2 Solusi.....	3
1.2.3 Pertanyaan Penelitian	3
1.3 Tujuan.....	3
1.4 Manfaat.....	4
1.5 Batasan Masalah.....	4
1.6 Metodologi Penulisan Skripsi	4
BAB II KAJIAN PUSTAKA	6
2.1 Berita	6
2.2 Natural Language Processing	6
2.3 Klasifikasi.....	7
2.4 Text Mining	7

2.5	Preprocessing.....	7
2.5.1	Normalisasi	7
2.5.2	Case Folding	8
2.5.3	Cleaning	8
2.5.4	Tokenizing.....	8
2.5.5	Stopword Removal.....	8
2.5.6	Stemming	8
2.6	TF-IDF.....	8
2.7	Deep Learning	9
2.8	Reccurent Neural Network (RNN).....	10
2.9	Long Short Term Memory	12
2.9.1	Forget Gate.....	13
2.9.2	Input Gate.....	14
2.9.3	Cell State	15
2.9.4	Output Gate	15
2.10	Layer LSTM	17
2.10.1	Input Layer	17
2.10.2	LSTM Units	17
2.10.3	Memory Cell	17
2.10.4	Hidden State	17
2.10.5	Output Layer	17
2.10.6	Dropout Layer	17
2.11	Fully connected layers.....	18
2.12	Fungsi Aktivasi.....	18
2.12.1	Sigmoid	18
2.12.2	Tanh.....	19

2.12.3	Softmax	20
2.13	Optimizer Adam	20
2.14	Batch Size, Epoch, Learning rate, dan binary cross-entropy	21
2.15	Neuron Hidden Layer	22
2.16	Confussion Matrix	22
2.16.1	Accuracy	23
2.16.2	Precision	23
2.16.3	Recall	24
2.16.4	F-Measure	24
2.17	Penelitian Terkait	24
BAB III METODE PENELITIAN		29
3.1	Desain Sistem	29
3.2	Dataset	30
3.3	Preprocessing	30
3.3.1	Normalisasi	31
3.3.2	Case Folding	31
3.3.3	Cleaning	31
3.3.4	Tokenizing	32
3.3.5	Stopword Removal	32
3.3.6	Stemming	33
3.4	TF-IDF	33
3.5	LSTM	36
3.5.1	Forget Gate	42
3.5.2	Input Gate	43
3.5.3	Output Gate	45
3.6	Evaluasi	47

3.7	Skenario Pengujian	47
BAB IV HASIL DAN PEMBAHASAN		50
4.1	Ruang Lingkup Kerja	50
4.2	Implementasi Sistem	50
4.2.1	Inisiasi <i>Dataset</i>	50
4.2.2	Memberikan Label pada Masing-Masing <i>Dataset</i>	51
4.2.4	Preprocessing	52
4.2.5	Pembobotan TF-IDF	57
4.2.6	Pembagian Data <i>Training</i> dan Data <i>Testing</i>	58
4.2.7	<i>Reshape</i> Data <i>Training</i> dan <i>Testing</i>	58
4.2.8	LSTM	59
4.2.9	Menjalankan Training pada Model LSTM	60
4.2.10	Melakukan prediksi pada model yang telah dibangun.....	61
4.2.11	Mencari nilai akurasi dari hasil prediksi	61
4.3	Implementasi Sistem	62
4.4	Lingkungan Uji Coba	64
4.5	Analisa Hasil	67
BAB V KESIMPULAN		71
5.1	Kesimpulan.....	71
5.2	Saran	71
REFERENSI		72
DAFTAR RIWAYAT HIDUP.....		75

DAFTAR GAMBAR

Gambar 2. 1. Arsitektur RNN	11
Gambar 2. 2. Arsitektur LSTM	13
Gambar 2. 3. Forget Gate	13
Gambar 2. 4. Input Gate	14
Gambar 2. 5. Cell State	15
Gambar 2. 6. Output Gate	16
Gambar 3. 1. Alur Penelitian.....	28
Gambar 3. 2. Perbandingan Jumlah dokumen berlabel Fake dan True	29
Gambar 3. 3. Alur TF-IDF	32
Gambar 3. 4. Alur Pemodelan LSTM dalam deteksi hoax	33
Gambar 3. 5. komponen utama yang ada di dalam LSTM Cell.....	33
Gambar 3. 6. Hasil perubahan representasi matriks TF-IDF menjadi Array pada tiap dokumen.....	35
Gambar 3. 7. penggabungan dua dokumen menjadi satu array 2 dimensi.....	35
Gambar 3. 8. array 2 dimensi	36
Gambar 3. 9. hasil perubahan array 2 dimensi menjadi array 3 dimensi	37
Gambar 3. 10. Visual Alur Forget Gate	40
Gambar 3. 11. Visual alur Input Gate	41
Gambar 3. 12. Visual alur new cell state	42
Gambar 3. 13. visual alur output gate	43
Gambar 3. 14. Alur Preprocessing pertama	45
Gambar 3. 15. Alur Preprocessing kedua.....	45
Gambar 4. 1. Tampilan Interface Sistem	63
Gambar 4. 2. Tampilan <i>interface</i> sistem apabila hasil deteksi berupa <i>non hoax</i> ..	63
Gambar 4. 3. Tampilan <i>interface</i> sistem apabila hasil deteksi berupa <i>non hoax</i> ..	64

DAFTAR TABEL

Tabel 2. 1 Penelitian Terkait	27
Tabel 3. 1. Contoh Normalisasi.....	31
Tabel 3. 2. Contoh <i>Casefolding</i>	31
Tabel 3. 3. Contoh <i>Cleaning</i>	32
Tabel 3. 4. Contoh <i>Tokenizing</i>	32
Tabel 3. 5. Contoh <i>Stopword Removal</i>	32
Tabel 3. 6. Contoh Stemming	33
Tabel 3. 7. Contoh hasil <i>term frequency</i>	34
Tabel 3. 8. Contoh perhitungan <i>inverse document frequency (IDF)</i>	34
Tabel 3. 9 Contoh perhitungan TF IDF	35
Tabel 3. 10. Representasi Matriks TF-IDF	37
Tabel 3. 11. nilai hasil TF-IDF dengan 4 fitur yang akan diolah dengan perhitungan LSTM.....	42
Tabel 3. 12. Hasil perhitungan LSTM	47
Tabel 4. 1 Kebutuhan Perangkat.....	50
Tabel 4. 2 Hasil Lingkungan Uji Coba	68

DAFTAR KODE PROGRAM

Kode Program 4. 1 Inisiasi Dataset.....	50
Kode Program 4.2 Pelabelan Masing-masing dataset.....	51
Kode Program 4. 3 menyatukan dataset.....	51
Kode Program 4.4 Case Folding	54
Kode Program 4. 5 Tokenizing	55
Kode Program 4. 6 Stopword Removal	56
Kode Program 4. 7 Stemming.....	57
Kode Program 4. 8 Pembobotan TF-IDF Pada dataset yang telah dilakukan Preprocessing	58
Kode Program 4. 9 Split Data menjadi Training dan Testing.....	58
Kode Program 4. 10 Reshape Data Training dan Testing.....	58
Kode Program 4. 11 Inisiasi Model LSTM yang akan dibangun	59
Kode Program 4. 12 perintah untuk menjalankan proses training LSTM	60