



SKRIPSI

IMPLEMENTASI *BIDIRECTIONAL*-LSTM DAN WORD2VEC DALAM KLASIFIKASI BERITA PALSU (HOAX) DI INDONESIA

FADHILAH NUR HIDAYAT
NPM 21083010082

DOSEN PEMBIMBING
Wahyu Syaifullah J. S, S.Kom., M.Kom.
Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

**KEMENTERIAN PENDIDIKAN, TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

IMPLEMENTASI *BIDIRECTIONAL*-LSTM DAN *WORD2VEC* DALAM KLASIFIKASI BERITA PALSU (*HOAX*) DI INDONESIA

Oleh:

FADHILAH NUR HIDAYAT
NPM. 21083010082

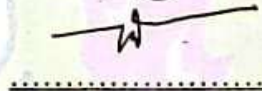
Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi
Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur
Pada tanggal 9 Januari 2026:

Wahyu Syaifullah J.S., S.Kom., M.Kom.
NIP. 19860825 202121 1 003



(Pembimbing I)

Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.
NIP. 19830310 202121 1 006



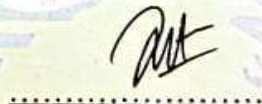
(Pembimbing II)

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009



(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 19990606 202406 1 001



(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

IMPLEMENTASI *BIDIRECTIONAL*-LSTM DAN *WORD2VEC* DALAM KLASIFIKASI BERITA PALSU (*HOAX*) DI INDONESIA

Oleh:

FADHILAH NUR HIDAYAT

NPM. 21083010082

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.

NIP. 198012052005011002

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : Fadhilah Nur Hidayat
NPM : 21083010082
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila di kemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 15 Januari 2026
Yang Membuat Pernyataan,



Fadhilah Nur Hidayat
NPM. 21083010082

ABSTRAK

Nama Mahasiswa / NPM : Fadhilah Nur Hidayat / 21083010082
Judul Skripsi : Implementasi *Bidirectional*-LSTM dan Word2Vec dalam Klasifikasi Berita Palsu (*Hoax*) di Indonesia
Dosen Pembimbing : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

Penyebaran berita palsu (*hoax*) di Indonesia semakin meningkat seiring perkembangan teknologi digital, yang berpotensi menimbulkan keresahan dan gangguan stabilitas di masyarakat. Penelitian ini bertujuan untuk membangun model klasifikasi berita palsu menggunakan pendekatan *Deep Learning* dengan algoritma *Bidirectional Long Short-Term Memory* (Bi-LSTM) dan *Word2Vec*. Data penelitian dikumpulkan melalui teknik *web scraping* dari portal Detik.com untuk berita valid dan Turnbackhoax.id untuk berita *hoax*, dengan total data yang digunakan setelah proses penyeimbangan sebanyak 2.800 artikel. Tahap pra-pemrosesan meliputi pembersihan teks, normalisasi, dan penghapusan *stopwords* dengan penanganan khusus pada kata negasi. Ekstraksi fitur dilakukan menggunakan *Word2Vec* arsitektur *Continuous Bag of Words* (CBOW) dengan dimensi 300 yang dilatih menggunakan korpus Wikipedia Bahasa Indonesia untuk menangkap makna semantik kata. Model Bi-LSTM digunakan untuk mempelajari pola urutan teks dari dua arah (maju dan mundur). Berdasarkan hasil evaluasi pada data uji (*testing set*), model yang diusulkan berhasil mencapai Akurasi sebesar 86,43%, dengan nilai Presisi 87,23%, *Recall* 85,36%, dan *F1-Score* 86,28% . Penelitian ini juga menghasilkan implementasi sistem antarmuka berbasis web yang mampu mendeteksi validitas berita secara otomatis.

Kata kunci : Berita Palsu (*Hoax*), *Deep Learning*, *Bidirectional* LSTM, Word2Vec, Klasifikasi Teks.

ABSTRACT

Student Name / NPM : Fadhilah Nur Hidayat / 21083010082
Thesis Title : *Implementation of Bidirectional-LSTM and Word2Vec for Fake News (Hoax) Classification in Indonesia*
Thesis Advisors : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom.
2. Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T.

The spread of fake news (hoax) in Indonesia is increasing with the development of digital technology, potentially causing public unrest and instability. This study aims to build a fake news classification model using a Deep Learning approach with the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm and Word2Vec. Research data was collected via web scraping from Detik.com for valid news and Turnbackhoax.id for hoaxes, with a total dataset of 2,800 articles used after the balancing process. The pre-processing stage includes text cleaning, normalization, and stopword removal with special handling for negation words. Feature extraction is performed using Word2Vec with Continuous Bag of Words (CBOW) architecture and 300 dimensions, pre-trained on the Indonesian Wikipedia corpus to capture semantic meaning. The Bi-LSTM model is utilized to learn text sequence patterns from two directions (forward and backward). Based on the evaluation of the testing set, the proposed model successfully achieved an Accuracy of 86.43%, with a Precision of 87.23%, Recall of 85.36%, and F1-Score of 86.28% . This research also produced a web-based interface system implementation capable of detecting news validity automatically.

Keywords: *Fake News (Hoax), Deep Learning, Bidirectional LSTM, Word2Vec, Text Classification*

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul “ **Implementasi *Bidirectional-LSTM* dan *Word2Vec* dalam Klasifikasi Berita Palsu (*HOAX*) di Indonesia**” dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Wahyu Syaifullah J. S, S.Kom., M.Kom. selaku Dosen Pembimbing utama dan Dr. Ir. Mohammad Idhom, S.P., S.Kom., M.T. selaku Dosen Pembimbing kedua yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Ayahanda Mulyono, beliau yang menjadi inti tulang punggung keluarga. Meskipun beliau tidak sempat merasakan pendidikan hingga bangku perkuliahan, namaun beliau mampu mendidik penulis hingga menjadi laki-laki yang kuat dan tegar dalam menghadapi segala rintangan, hingga penulis mampu menyelesaikan studinya hingga sarjana.
2. Ibunda Sumarni, beliau sangat berperan penting dalam terselesainya program studi penulis. Beliau juga tidak sempat merasakan pendidikan hingga bangku perkuliahan, namun senantiasa gigih dalam memanjatkan doa meminta kepada Tuhan Yang Maha Esa, hingga penulis mampu menyelesaikan studinya hingga sarjana.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Ibu Kartika Maulida Hindrayani S.Kom, M.Kom. selaku dosen wali yang selalu memberikan bimbingan dan segala bentuk bantuan akademik yang dibutuhkan penulis selama masa studi di Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.

6. Segenap dosen-dosen Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur.
7. Dan yang terakhir untuk diri penulis sendiri Fadhilah Nur Hidayat yang telah bekerja keras dan bertahan sehingga dapat bertanggung jawab untuk menyelesaikan apa yang telah penulis mulai.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 15 Januari 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xiii
DAFTAR LAMPIRAN	xiv
DAFTAR NOTASI.....	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu	7
2.2 Kerangka Teori	10
2.2.1. <i>Media Online</i>	10
2.2.2. Berita Palsu (<i>Hoax</i>).....	11
2.2.3. Detik.com.....	12
2.2.4. <i>Turnbackhoax.id</i>	13
2.2.5. <i>Artificial Intelligence</i>	14
2.2.6. <i>Machine Learning</i>	15
2.2.7. <i>Deep Learning</i>	16
2.2.8. <i>Artificial Neural Network</i>	18

2.2.9.	<i>Recurrent Neural Network</i>	19
2.2.10.	<i>Long Short-Term Memory</i>	22
2.2.11.	<i>Bidirectional Long Short-Term Memory</i>	24
2.2.12.	<i>Word Embedding</i>	24
2.2.13.	<i>K-Fold Cross Validation</i>	27
2.2.14.	Evaluasi Model.....	29
BAB III METODOLOGI PENELITIAN.....		33
3.1.	Variabel Penelitian dan Sumber Data.....	33
3.2.	Langkah Analisis	42
3.3.	Desain Sistem	55
BAB IV HASIL DAN PEMBAHASAN.....		57
4.1	Pengumpulan Data.....	57
4.2	<i>Exploratory Data Analysis (EDA)</i>	62
4.3	<i>Preprocessing Data</i>	67
4.4	Pembagian Dataset	75
4.5	Ekstraksi Fitur dengan <i>Word2Vec (Feature Extraction)</i>	77
4.6	Konversi Data ke Sekuens (<i>Sequences & Padding</i>).....	83
4.7	Implementasi Model BiLSTM.....	84
4.8	Pelatihan dengan <i>K-Fold Cross Validation</i>	86
4.9	Evaluasi pada Data Testing (<i>Unseen Data</i>).....	91
4.10	Analisis Hasil dengan Perluasan Dataset.....	94
4.11	Analisis Komparatif Dampak Penambahan Data Dataset	98
4.12	Pengujian Sistem Prediksi (<i>System Implementation</i>)	99
4.13	<i>Graphical User Interface (GUI)</i>	101
BAB V PENUTUP		117
5.1.	Kesimpulan.....	117
5.2.	Saran Pengembangan.....	118
DAFTAR PUSTAKA		121
LAMPIRAN		127

DAFTAR GAMBAR

Gambar 2. 1 Website Berita Detik.com	13
Gambar 2. 2 Website Berita Turnbackhoax.id	14
Gambar 2.3 Penjelasan Jenis <i>Machine Learning</i>	15
Gambar 2. 4 Alur Kerja <i>Deep Learning</i>	17
Gambar 2. 5 Alur Kerja ANN	18
Gambar 2. 6 Alur Kerja RNN.....	20
Gambar 2. 7 Alur Gate LSTM.....	22
Gambar 2. 8 Alur Kerja BiLSTM.....	24
Gambar 2. 9 Alur Kerja <i>Word2Vec</i>	25
Gambar 2.10 Alur Kerja <i>Skip-Gram</i>	26
Gambar 2. 11 Alur Kerja <i>Continuous Bag of Words</i> (CBOW).....	26
Gambar 2. 12 Proses Kerja <i>K-Fold Cross Validation</i>	28
Gambar 3. 1 Pelabelan Berita	42
Gambar 3. 2 Alur Penelitian	43
Gambar 3. 3 Tahapan Pemrosesan Data	45
Gambar 3. 4 Diagram Alir Pra-Pelatihan Word2Vec	49
Gambar 3. 5 Arsitektur Model Bi-LSTM.....	52
Gambar 3. 6 Desain Sistem Interface	55
Gambar 4. 1 Hasil Proses <i>Bias Marker Removal</i>	69
Gambar 4. 2 Hasil Normalisasi Data Berita	70
Gambar 4. 3 Hasil Stopwords Kata dalam Berita.....	71
Gambar 4. 4 Visualisasi <i>WordCloud</i> Berita Keseluruhan	72
Gambar 4. 5 Visualisasi <i>WordCloud</i> Berita <i>HOAX</i>	72
Gambar 4. 6 Visualisasi <i>WordCloud</i> Berita <i>Valid</i>	73
Gambar 4. 7 Visualisasi Distribusi Data Berita.....	75
Gambar 4. 8 Hasil <i>Testing Similarity</i> Dengan Model <i>Word2Vec</i>	81
Gambar 4. 9 Visualisasi <i>Clustering PCA</i>	82
Gambar 4. 10 Grafik <i>Training Loss</i> dan <i>Validation Loss</i>	89
Gambar 4. 11 Grafik <i>Training Accuracy</i> dan <i>Validation Accuracy</i>	90
Gambar 4. 12 Grafik <i>Precision–Recall Curve</i>	91

Gambar 4. 13	Visualisasi <i>Confusion Matrix</i>	94
Gambar 4. 14	Grafik <i>Training Loss</i> dan <i>Validation Loss</i> Penambahan Data.....	95
Gambar 4. 15	Grafik <i>Training Accuracy</i> dan <i>Validation Accuracy</i> Penambahan Data.....	96
Gambar 4. 16	Grafik <i>Precision–Recall Curve</i> Penambahan Data	96
Gambar 4. 17	Visualisasi <i>Confusion Matrix</i> Penambahan Data.....	97
Gambar 4. 18	Hasil Pengujian Model.....	100
Gambar 4. 19	Halaman Awal Sistem.....	113
Gambar 4. 20	Pengecekan Kalimat Berita	114
Gambar 4. 21	Hasil Prediksi Kalimat Berita	114
Gambar 4. 22	Hasil Analisis Kalimat Prediksi Berita	115

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	7
Tabel 2. 2 Tabel Kebenaran	29
Tabel 3. 1 Data Awal Berita Detik.com	33
Tabel 3. 2 Data Awal Berita Turnbackhoax.id.....	37
Tabel 3. 3 Pelabelan Berita.....	41
Tabel 4. 1 Karakteristik Data Yang Digunakan	57
Tabel 4. 2 Hasil Kata Teratas Berdasarkan <i>Bigram</i>	73
Tabel 4. 3 Struktur Parameter Model Yang Digunakan	78
Tabel 4. 4 Hasil Sekuens Pada Data Berita	83
Tabel 4. 5 Struktur <i>Layer</i> Pada <i>Bi-LSTM</i>	84
Tabel 4. 6 Total Parameter Model <i>Bi-LSTM</i>	86
Tabel 4. 7 Hasil Training Pada Tiap Fold	88
Tabel 4. 8 Hasil Uji Metrik Evaluasi.....	92
Tabel 4. 9 Hasil Pelatihan Penambahan Data.....	95
Tabel 4. 10 Hasil Uji Metrik Evaluasi Penambahan Data.....	97
Tabel 4. 11 Perbandingan Performa Model pada Data <i>Testing</i>	98

DAFTAR LAMPIRAN

Lampiran 1. Dataset	127
Lampiran 2. Aplikasi GUI Streamlit	128
Lampiran 3. <i>Letter of Accpetance</i> (LoA)	129

DAFTAR NOTASI

I	:	Intensitas
W_R	:	<i>weight factor</i>
H	:	<i>hue</i>
S	:	<i>saturation</i>
V	:	<i>value</i>
dst	:	Gambar akumulator
scr	:	Gambar Input
F	:	<i>Foreground</i>
B	:	<i>Background</i>
f	:	<i>Frame</i>
SE	:	<i>Structuring Element</i>
ECD	:	<i>Equivalent Circular Diameter</i>
b	:	bias
$WED(f_i \rightarrow , prototype)$	:	bobot <i>euclidian distance</i> antara vektor fitur $f_i \rightarrow$
$f(x)$	:	Fungsi Vektor Masukan
$d(x', x)$	:	jarak di antara data uji z ke setiap vector data latih
$K(x, y)$	:	fungsi kernel linear
$G(x)$	:	fungsi Gaussian satu dimensi
σ	:	standard deviasi dari distribusi
$G(x, y)$	:	fungsi Gaussian dua dimensi