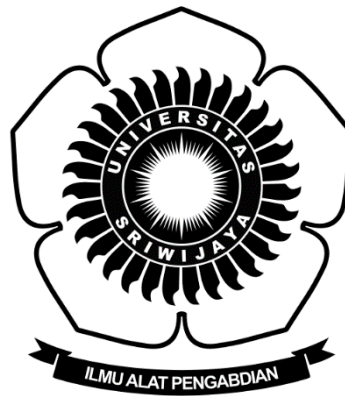


KLASIFIKASI BERITA HOAX BAHASA INDONESIA MENGUNAKAN LSTM (LONG SHORT TERM MEMORY)

Diajukan Sebagai Syarat Untuk
Menyelesaikan Pendidikan Program Strata-1
Pada Jurusan Teknik Informatika



Oleh :

Faiq Fadlurrahman
NIM : 09021382025119

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

KLASIFIKASI BERITA HOAX BAHASA INDONESIA
MENGUNAKAN LSTM (LONG SHORT TERM MEMORY)

Oleh :

Faiq Fadlurrahman

NIM : 09021382025119

Palembang, 15 Maret 2024

Pembimbing I



Alvi Syahrini Utami, M.Kom.
NIP 197812222006042003

Pembimbing II



Junia Kurniati, M.Kom.
NIP 198906262024212001

Mengetahui,

Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.
NIP 198005222008121002

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari ini tanggal 15 Maret 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Faiq Fadlurrahman

NIM : 09021382025119

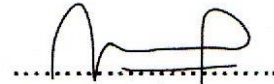
Judul : Klasifikasi Berita Hoax Bahasa Indonesia Menggunakan LSTM (Long Short Term Memory)

dan dinyatakan LULUS.

1. Ketua Penguji

Alfarissi, M. Cs.

NIP 198512152014041001



2. Pembimbing I

Alvi Syahrini, M.Kom.

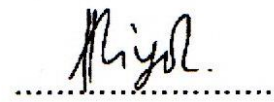
NIP 197812222006042003



3. Pembimbing II

Junia Kurniati, M.Kom.

NIP 198906262024212001



4. Penguji

Desty Rodiah, M.T.

NIP 198912212020122011



Mengetahui,
Ketua Jurusan Teknik Informatika



Dr. M. Fachrurrozi, S.Si., M.T.
NIP 198005222008121002

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Faiq Fadlurrahman

NIM : 09021382025119

Program Studi : Teknik Informatika Bilingual

Judul : Klasifikasi Berita Hoax Bahasa Indonesia Menggunakan LSTM
(Long Short Term Memory)

Hasil Pengecekan Software *iThenticate/Turnitin*: 19%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

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



Palembang, 26 Maret 2024



Faiq Fadlurrahman

NIM 09021382025119

MOTTO DAN PERSEMBAHAN

“Never Give Up”

Kupersembahkan karya tulis ini kepada:

- Orang Tua, Saudari, dan Keluargaku
- Dosen Pembimbing
- Teman-teman seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Hoax news poses a serious threat in the current digital environment as it harms society by disseminating false and misleading information. Detecting hoax news is essential to maintain the integrity of information, prevent the spread of misinformation, and safeguard public safety. This research adopts a classification approach to hoax news based on LSTM (Long Short-Term Memory) with the utilization of Word2Vec for word representation. This approach is chosen because LSTM can comprehend the context and sequence of words in a sentence, while Word2Vec provides a meaningful representation of words. By combining the strengths of both methods, this study aims to contribute to improving the detection of Indonesian hoax news. The data used in this study consist of 17,656 records, with 2 classes namely real and hoax. The research findings indicate that the LSTM model with the best configuration, including a learning rate of 0.0001, dropout of 0.2, hidden layer of 64, and batch size of 64, achieves high accuracy in classifying hoax news. The model evaluation using accuracy metric of 94%, recall of 96%, precision of 93%, and f1-score of 94% confirms the reliability of this model in identifying and distinguishing hoax news with a high level of accuracy.

Key word : Long Short Term Memory, Word2Vec, Hoax News, News Classification

ABSTRAK

Berita *hoax* menjadi ancaman serius dalam lingkungan digital saat ini karena merugikan masyarakat dengan menyebarkan informasi yang tidak benar dan menyesatkan. Deteksi berita *hoax* menjadi esensial untuk menjaga integritas informasi, mencegah penyebaran hoaks, dan menjaga keamanan publik. Penelitian ini menggunakan pendekatan klasifikasi berita *hoax* berbasis LSTM (Long Short-Term Memory) dengan penggunaan *Word2Vec* untuk representasi kata. Pendekatan ini dipilih karena LSTM mampu memahami konteks dan urutan kata dalam kalimat, sementara *Word2Vec* memberikan representasi kata yang kaya makna. Dengan menggabungkan keunggulan kedua metode ini, penelitian ini bertujuan untuk memberikan kontribusi dalam meningkatkan deteksi berita *hoax* bahasa Indonesia. Data yang digunakan dalam penelitian ini berjumlah 17,656 data dimana terdapat 2 kelas yaitu *real* dan *hoax*. Hasil penelitian menunjukkan bahwa model LSTM dengan konfigurasi terbaik, yaitu *learning rate* 0,0001, *dropout* 0,2, *hidden layer* 64, dan *batch size* 64, mencapai akurasi tinggi dalam mengklasifikasikan berita *hoax*. Evaluasi model menggunakan metrik akurasi 94%, *recall* 96%, *precision* 93%, dan *f1-score* 94% menegaskan kehandalan model ini dalam mengenali dan memisahkan berita *hoax* dengan tingkat akurasi yang tinggi.

Kata kunci : *Long Short Term Memory*, *Word2Vec*, Berita *Hoax*, Klasifikasi Berita

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Allah SWT, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan penulisan skripsi ini dengan judul "**Klasifikasi Berita Hoax Bahasa Indonesia Menggunakan LSTM (Long Short Term Memory)**". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Universitas Sriwijaya.

Penulisan skripsi ini merupakan perjalanan panjang yang penuh liku-liku, yang tidak akan terlaksana tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Keluarga, yang selalu memberikan doa, dukungan, dan motivasi dalam setiap langkah perjalanan ini.
2. Pembimbing Skripsi, Ibu Alvi Syahrini Utami, M.Kom., dan Ibu Junia Kurniati, M.Kom, yang telah memberikan bimbingan, arahan, dan masukan berharga dari awal hingga akhir penulisan skripsi ini.
3. Dosen-dosen dan staf Program Studi Teknik Informatika, yang telah memberikan ilmu pengetahuan dan pengalaman berharga selama studi di universitas.
4. Mba Wiwin dan Mba Rika selaku admin Jurusan Teknik Informatika yang telah membantu mengurus berkas administrasi penulis.
5. Teman-teman seperjuangan, yang selalu memberikan semangat, dukungan, dan kerjasama dalam menyelesaikan tugas akhir ini.
6. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah turut serta memberikan kontribusi dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan guna perbaikan di masa mendatang. Semoga skripsi ini dapat bermanfaat bagi pengembangan ilmu pengetahuan dan teknologi di masa yang akan datang.

Akhir kata, penulis berharap agar skripsi ini dapat menjadi sumbangan kecil dalam perjalanan ilmiah dan pengembangan ilmu pengetahuan di bidang Teknik Informatika.

Palembang, 26 Maret 2024

A handwritten signature in black ink, appearing to be 'Faiq' with a stylized flourish below it.

Faiq Fadlurrahman

DAFTAR ISI

| | Halaman |
|--|---------|
| HALAMAN JUDUL..... | i |
| LEMBAR PENGESAHAN SKRIPSI..... | ii |
| TANDA LULUS UJIAN KOMPREHENSIF..... | iii |
| HALAMAN PERNYATAAN | iv |
| MOTTO DAN PERSEMBAHAN | v |
| ABSTRACT | vi |
| ABSTRAK | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | x |
| DAFTAR TABEL..... | xiii |
| DAFTAR GAMBAR | xv |
| DAFTAR ISTILAH, SINGKATAN, DAN LAMBANG | xvi |
| | |
| BAB I PENDAHULUAN | |
| 1.1 Pendahuluan | I-1 |
| 1.2 Latar Belakang | I-1 |
| 1.3 Rumusan Masalah | I-5 |
| 1.4 Tujuan Penelitian | I-6 |
| 1.5 Manfaat Penelitian | I-6 |
| 1.6 Batasan Masalah..... | I-6 |
| 1.7 Sistematika Penulisan | I-7 |
| 1.8 Kesimpulan | I-8 |
| | |
| BAB II KAJIAN LITERATUR | |
| 2.1 Pendahuluan | II-1 |
| 2.2 Landasan Teori..... | II-1 |
| 2.2.1 <i>Natural Language Processing</i> | II-1 |
| 2.2.2 Klasifikasi Teks | II-1 |
| 2.2.3 <i>Word Embedding</i> | II-2 |

| | |
|---|-------|
| 2.2.4 <i>Long Short Term Memory</i> | II-3 |
| 2.2.5 <i>Confusion Matrix</i> | II-8 |
| 2.2.6 <i>Rational Unified Process (RUP)</i> | II-10 |
| 2.3 Penelitian Lain yang Relevan..... | II-13 |
| 2.4 Kesimpulan | II-15 |

BAB III METODOLOGI PENELITIAN

| | |
|--|--------|
| 3.1 Pendahuluan | III-1 |
| 3.2 Pengumpulan Data | III-1 |
| 3.3 Tahapan Penelitian | III-4 |
| 3.3.1 Kerangka Kerja Penelitian..... | III-5 |
| 3.3.2 Alat yang Digunakan dalam Penelitian | III-20 |
| 3.3.3 Menentukan Kriteria Pengujian..... | III-21 |
| 3.4 Metode Pengembangan Perangkat Lunak..... | III-22 |
| 3.4.1 Fase Insepsi..... | III-22 |
| 3.4.2 Fase Elaborasi..... | III-22 |
| 3.4.3 Fase Konstruksi | III-23 |
| 3.4.4 Fase Transisi | III-23 |
| 3.5 Kesimpulan | III-23 |

BAB IV PENGEMBANGAN PERANGKAT LUNAK

| | |
|--|------|
| 4.1 Pendahuluan | IV-1 |
| 4.2 Fase Insepsi | IV-1 |
| 4.2.1 Pemodelan Bisnis | IV-1 |
| 4.2.2 Kebutuhan Sistem..... | IV-2 |
| 4.2.3 Analisis dan Desain | IV-3 |
| 4.3 Fase Elaborasi | IV-6 |
| 4.3.1 Perancangan Data | IV-6 |
| 4.3.2 Rancangan <i>Interface</i> | IV-6 |
| 4.3.3 Kebutuhan Sistem..... | IV-7 |
| 4.3.4 Diagram <i>Activity</i> | IV-7 |
| 4.3.5 <i>Sequence</i> Diagram | IV-8 |
| 4.4 Fase Konstruksi..... | IV-9 |

| | |
|---|-------|
| 4.4.1 <i>Class Diagram</i> | IV-10 |
| 4.4.2 Implementasi Kelas | IV-10 |
| 4.4.3 Implementasi <i>Interface</i> | IV-11 |
| 4.5 Fase Transisi | IV-12 |
| 4.5.1 Rencana Pengujian <i>Use Case</i> | IV-12 |
| 4.5.2 Pengujian <i>Use Case</i> | IV-12 |
| 4.6 Kesimpulan | IV-13 |

BAB V HASIL DAN PEMBAHASAN

| | |
|---|------|
| 5.1 Pendahuluan | V-1 |
| 5.2 Data Hasil Penelitian..... | V-1 |
| 5.2.1 Konfigurasi Hasil Percobaan | V-1 |
| 5.2.2 Data Hasil Konfigurasi | V-4 |
| 5.2.3 Analisis Hasil Penelitian..... | V-9 |
| 5.3 Kesimpulan | V-17 |

BAB V KESIMPULAN DAN SARAN

| | |
|-----------------------|------|
| 6.1 Pendahuluan | VI-1 |
| 6.2 Kesimpulan | VI-1 |
| 6.3 Saran..... | VI-2 |

| | |
|----------------------|-----|
| DAFTAR PUSTAKA | xix |
|----------------------|-----|

| | |
|---------------|-----|
| LAMPIRAN..... | xxi |
|---------------|-----|

DAFTAR TABEL

| | Halaman |
|---|---------|
| II-1. Tabel <i>Confusion Matrix</i> | II-9 |
| III-1. Tabel Contoh Data Berita..... | III-2 |
| III-2. Tabel Contoh <i>Preprocessing</i> | III-7 |
| III-3. Tabel Judul Berita | III-9 |
| III-4. Tabel Hasil Proses <i>Case Folding</i> | III-9 |
| III-5. Tabel Hasil Proses <i>Cleaning</i> | III-10 |
| III-6. Tabel Hasil <i>Stopword Removal</i> | III-11 |
| III-7. Tabel Kamus Token | III-11 |
| III-8. Tabel Hasil Proses Tokenisasi | III-12 |
| III-9. Tabel Hasil Proses <i>Padding</i> | III-13 |
| III-10. Tabel Contoh <i>Word Vector</i> 50 Dimensi..... | III-14 |
| III-11. Tabel Contoh <i>Embedding</i> | III-15 |
| III-12. Tabel Contoh Bobot dan <i>Bias</i> | III-15 |
| III-13. Tabel Rancangan Hasil <i>Training</i> | III-21 |
| III-14. Tabel Rancangan Hasil <i>Confusion Matrix</i> | III-22 |
| III-15. Tabel Rancangan Hasil Pengujian | III-22 |
| IV-1. Tabel Kebutuhan Fungsional Perangkat Lunak | IV-2 |
| IV-2. Tabel Kebutuhan Non-Fungsional Perangkat Lunak..... | IV-2 |
| IV-3. Tabel Definisi Aktor | IV-4 |
| IV-4. Tabel Definisi <i>Use Case</i> | IV-4 |
| IV-5. Tabel Skenario <i>Use Case</i> Klasifikasi Berita | IV-5 |
| IV-6. Tabel Implementasi Kelas..... | IV-11 |
| IV-7. Tabel Rencana Pengujian <i>Use Case</i> | IV-12 |
| IV-8. Tabel Pengujian <i>Use Case</i> | IV-13 |
| V-1. Tabel Pengujian <i>Use Case</i> | V-2 |
| V-2. Tabel Hasil <i>Training</i> Tiap Model | V-4 |
| V-3. Tabel <i>Confusion Matrix</i> Setiap Model..... | V-10 |
| V-4. Tabel <i>Confusion Matrix</i> Model 3 | V-11 |
| V-5. Tabel <i>Classification Report</i> Setiap Model | V-13 |

| | | |
|------|---|------|
| V-6. | Tabel Hasil <i>Confusion Matrix</i> Model 3 | V-15 |
| V-7. | Tabel Hasil Data Uji Model 3 | V-15 |
| V-8. | Tabel Contoh Klasifikasi Berita..... | V-16 |

DAFTAR GAMBAR

| | Halaman |
|--|---------|
| II-1. Gambar Arsitektur LSTM | II-4 |
| III-1. Gambar Distribusi Jumlah Data per Kategori | III-2 |
| III-2. Gambar Diagram Tahapan Penelitian | III-4 |
| III-3. Gambar Diagram Kerangka Kerja | III-5 |
| IV-1. Gambar <i>Use Case</i> Diagram..... | IV-3 |
| IV-2. Gambar Rancangan <i>Interface</i> | IV-6 |
| IV-3. Gambar Diagram <i>Activity</i> | IV-8 |
| IV-4. Gambar <i>Sequence</i> Diagram..... | IV-9 |
| IV-5. Gambar <i>Class</i> Diagram | IV-10 |
| IV-6. Gambar Implementasi <i>Interface</i> | IV-11 |
| V-1. Gambar Akurasi <i>Training</i> dan <i>Testing</i> Model 3 | V-6 |
| V-2. Gambar <i>Loss Training</i> dan <i>Testing</i> Model 3 | V-7 |
| V-3. Gambar <i>Loss Training</i> dan <i>Testing</i> Model | V-9 |
| V-4. Gambar <i>Confusion Matrix</i> Model 3 | V-14 |

DAFTAR ISTILAH , SINGKATAN, DAN LAMBANG

| | |
|---------------------------|---|
| <i>Bias</i> | : Dalam konteks jaringan saraf, <i>bias</i> adalah parameter yang ditambahkan ke keluaran dari setiap unit atau <i>neuron</i> dalam lapisan. |
| CNN | : <i>Convolutional Neural Network</i> |
| <i>Dropout</i> | : Teknik regularisasi yang digunakan dalam jaringan saraf. Ini berfungsi dengan mengatur secara acak sebagian unit (<i>neuron</i>) dalam jaringan saraf untuk dinonaktifkan (<i>dropout</i>) selama pelatihan |
| <i>Early Stopping</i> | : Teknik dalam <i>machine learning</i> untuk menghentikan proses pelatihan model sebelum mencapai titik <i>overfitting</i> . |
| <i>Exploding Gradient</i> | : <i>Exploding gradient</i> terjadi saat nilai-nilai gradien (<i>gradient</i>) yang digunakan untuk mengoptimalkan model tumbuh sangat cepat, sehingga dapat mengakibatkan nilai-nilai parameter dalam model menjadi sangat besar |
| <i>Forget Gate</i> | : <i>Gate</i> yang bertugas untuk melupakan beberapa informasi yang tidak relevan dan sudah tidak diperlukan oleh sebuah sistem |
| <i>Input Gate</i> | : <i>Gate</i> yang bertugas untuk memasukkan informasi yang berguna untuk mendukung keakuratan data |
| <i>Output Gate</i> | : <i>Gate</i> yang menjadi gerbang terakhir untuk menghasilkan informasi data yang komplet dan aktual |
| <i>Hidden Layer</i> | : Lapisan di antara lapisan <i>input</i> dan lapisan <i>output</i> dalam jaringan saraf. Dalam konteks LSTM, tempat model LSTM menyimpan informasi kontekstual dalam data sekuensial |
| <i>Hidden Unit</i> | : Unit-unit atau sel-sel yang ada dalam lapisan tersembunyi jaringan LSTM |
| LSTM | : <i>Long Short Term Memory</i> |
| <i>Overfitting</i> | : <i>Overfitting</i> terjadi ketika model <i>machine learning</i> , termasuk jaringan saraf, belajar terlalu baik pada data pelatihan hingga tidak dapat umumkan hasilnya pada data baru atau uji. |
| SVM | : <i>Support Vector Machine</i> |
| <i>Vanishing Gradient</i> | : Gradien tumbuh sangat kecil atau mendekati nol selama proses pelatihan |

1D CNN

: 1 *Dimension Convolutional Neural Network*

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini akan menjelaskan tentang latar belakang penelitian, permasalahan yang akan diteliti, tujuan yang ingin dicapai, manfaat dari penelitian, serta batasan masalah. Selain itu akan diberikan gambaran umum mengenai bagaimana penelitian ini akan disusun dan kesimpulan dari bab ini.

1.2 Latar Belakang

Menurut Kamus Besar Bahasa Indonesia (KBBI), hoaks berarti berita bohong (Wijayanti, 2019). *Hoax* merupakan berita yang masih diragukan keaslian atau kebenarannya dan membuat masyarakat sangat resah karena informasi tersebut belum dapat dipastikan kebenarannya (Fauzi et al., 2019). Pada tahun 2016, *Central Connecticut University* merilis penelitian bertajuk *World's Most Literate Nations* dimana Indonesia menduduki peringkat 60 dari 61 negara yang berpartisipasi. Hal ini menunjukkan bahwa tingkat literasi media di Indonesia, kemampuan mengevaluasi informasi secara kritis, dan membedakan berita palsu dengan berita asli masih kurang.

Hasil survei yang dilakukan oleh Katadata Insight Center (KIC) yang bekerja sama dengan Kementerian Komunikasi dan Informatika mengindikasikan bahwa setidaknya 30% sampai 60% orang Indonesia terpapar *hoax* saat mengakses dan berinteraksi melalui dunia maya. Sementara hanya 21% hingga 36% orang

yang dapat membedakan hoaks, sebagian besar terkait dengan masalah agama, politik, dan kesehatan¹).

Metode *machine learning* dan *deep learning* telah mengalami perkembangan yang pesat dalam beberapa waktu terakhir. Metode *deep learning* secara bertahap menunjukkan kinerja yang baik di banyak aplikasi, seperti analisis sentimen dan klasifikasi teks. *Deep learning* adalah pilihan yang lebih baik untuk *Natural Language Processing* dibandingkan dengan metode tradisional (Wang et al., 2018). Hal ini didukung juga oleh penelitian yang dilakukan oleh Chen et al., (2021) yang menyatakan bahwa performa *deep learning* untuk klasifikasi teks lebih baik dibandingkan dengan metode *machine learning* tradisional (Chen et al., 2021).

Dalam penelitian yang dilakukan oleh Nayoga et al., (2021) dengan judul “*Hoax Analyzer for Indonesian News Using Deep Learning Models*”, yang membandingkan beberapa metode *deep learning* seperti LSTM, CNN 1D, SVM, dan Naïve Bayes pada klasifikasi teks. Penelitian ini dilakukan untuk meningkatkan kredibilitas informasi di era digital, khususnya dalam deteksi berita palsu dalam bahasa Indonesia, yang memiliki karakteristik unik. Ditemukan bahwa model CNN 1D memiliki nilai *precision*, *recall*, dan *F1-score* sebesar 97% yang lebih tinggi dibandingkan dengan model *machine learning* tradisional (Nayoga et al., 2021).

Dalam penelitian yang pernah dilakukan oleh Bahad et al., (2019) dalam jurnalnya dengan judul “*Fake News Detection using Bi-directional LSTM-Recurrent Neural Network*”, yang menggunakan metode Bi-LSTM untuk

¹ Surat kabar online “Berita Satu”, 20 November 2020

mengklasifikasikan berita. Penelitian ini dilakukan untuk mengatasi masalah penyebaran berita *hoax* di media sosial dan menyediakan kerangka kerja untuk deteksi berita *hoax* yang efektif. Peneliti menggunakan dataset yang berasal dari *Kaggle* dan menggunakan fungsi optimasi *RMSProp* dan *Adam*. Hasil pada model CNN memiliki nilai akurasi *train* 99%, akurasi validasi dan akurasi tes 98% (Bahad et al., 2019).

Studi lain mengenai klasifikasi teks dilakukan oleh Sari et al., (2020) dalam jurnalnya yang berjudul “*Multilabel Classification for News Article Using Long Short-Term Memory*”. Hasil penelitian menunjukkan bahwa model mencapai tingkat akurasi sekitar 93,15%. Selain itu, nilai-nilai lainnya seperti *precision*, *recall*, dan *F1-score* juga sangat baik, yakni mencapai 94% (Sari et al., 2020).

Kemudian penelitian lain terkait klasifikasi teks telah dilakukan oleh Arief et al., (2023) dalam jurnal dengan judul “Penerapan Metode *Long Short Term Memory* Untuk Klasifikasi Pada *Hate Speech*”. Penelitian ini dilakukan untuk mengatasi masalah *hate speech* di media sosial dan membantu dalam pengembangan metode klasifikasi yang lebih canggih menggunakan LSTM. Dataset yang digunakan dalam penelitian ini berasal dari *Kaggle* dan terdiri dari 13,170 data dengan 12 kategori yang berbeda. Model memiliki performa terbaik dengan akurasi yang diperoleh pada data validasi sebesar 87,10% (Arief et al., 2023).

LSTM adalah salah satu modifikasi layer dari *Recurrent Neural Network* (RNN). LSTM bisa menyimpan informasi yang diperlukan dan menghilangkan informasi yang tidak dibutuhkan. Sel LSTM berfungsi untuk mempertahankan

informasi jangka panjang, sementara *hidden state* digunakan untuk menyimpan informasi jangka pendek. LSTM memiliki keunggulan yang jelas dibandingkan RNN karena mengatasi masalah *exploding* dan *vanishing gradient* (Kumar and Kishor, 2019). Unit LSTM telah berhasil diterapkan dalam pembelajaran *sequence* dan digunakan untuk memahami konteks serta struktur dalam konteks pemrosesan bahasa alami (NLP) (Srivastava et al., 2015). Berbeda dengan unit rekursif tradisional, unit LSTM mengatur memori pada setiap langkahnya, tanpa menggantikannya. Hal ini membuatnya lebih efektif dalam menangani ketergantungan jangka panjang dan menemukan fitur-fitur penting dalam rangkaian kalimat (Bengio et al., 1994).

Word2Vec adalah teknik populer yang digunakan dalam Pemrosesan Bahasa Alami (PBA) untuk *embedding* kata, yang merupakan proses merepresentasikan kata sebagai vektor padat dalam ruang vektor kontinu. Teknik ini diperkenalkan oleh para peneliti di Google pada tahun 2013 dan sejak itu telah banyak digunakan dalam berbagai tugas NLP karena efektivitasnya dalam menangkap hubungan semantik antara kata-kata. *Embedding* ini kemudian dapat digunakan sebagai fitur untuk tugas NLP *downstream* seperti analisis sentimen, klasifikasi teks, dan pengenalan entitas bernama. Secara keseluruhan, *Word2Vec* adalah alat yang sangat berguna dalam NLP yang membantu menangkap makna semantik kata-kata dalam ruang vektor berdimensi tinggi, memungkinkan representasi dan pemahaman yang lebih baik tentang data teks.

Pilihan metode LSTM dipilih untuk klasifikasi berita *hoax* karena kemampuannya menangani data berurutan, seperti teks berita, dengan menangkap

pola dan dependensi jarak jauh. Keunggulan LSTM dalam mengatasi masalah *vanishing gradient* membuatnya sesuai untuk tugas klasifikasi teks yang kompleks.

Dalam proses klasifikasi berita *hoax* dengan metode LSTM, langkah-langkah melibatkan tahap *preprocessing* teks untuk membersihkan dan mengubah teks menjadi format yang sesuai untuk pembuatan model. Selanjutnya, model LSTM dibangun menggunakan teks yang telah diproses sebagai *input*. Pelatihan model dilakukan dengan dataset yang berisi label berita *hoax* dan *real*. Setelah pelatihan selesai, model diuji pada data uji untuk mengukur kinerjanya dalam mengklasifikasikan berita sebagai *hoax* atau *real*.

Meskipun penelitian sebelumnya telah mencapai akurasi tinggi, penelitian ini dilakukan karena setiap dataset dan konteks penelitian dapat berbeda. Alasan melibatkan upaya untuk menggeneralisasi model, menghadapi perubahan konteks dan lingkungan informasi, serta berusaha meningkatkan kinerja model yang sudah ada atau mengeksplorasi variasi konsep dan pendekatan. Dengan demikian, penelitian ini berkontribusi pada pemahaman lebih lanjut tentang penggunaan metode LSTM dalam konteks klasifikasi berita *hoax* yang spesifik.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah yang akan diteliti adalah sebagai berikut:

1. Bagaimana mengimplementasikan model LSTM untuk mengklasifikasi berita *hoax*?

2. Bagaimana kinerja LSTM untuk klasifikasi berita *hoax* bahasa Indonesia?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan penelitian ini adalah:

1. Menghasilkan perangkat lunak untuk klasifikasi berita *hoax* bahasa Indonesia.
2. Mengetahui kinerja LSTM untuk klasifikasi berita *hoax* bahasa Indonesia.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah:

1. Penelitian ini dapat menjadi referensi tentang penggunaan LSTM dalam klasifikasi berita *hoax* bahasa Indonesia bagi akademisi pada khususnya.
2. Hasil penelitian ini dapat membantu dalam mendeteksi berita *hoax* yang beredar di Indonesia, sehingga dapat mencegah tersebarnya berita *hoax* bagi masyarakat Indonesia pada umumnya.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah:

1. Data yang digunakan adalah data judul berita.
2. Hanya terdiri dari dua label yaitu *Real(0)* dan *Hoax(1)*.
3. Data judul berita dominan kategori politik.

1.7 Sistematika Penulisan

Sistematika penulisan mengikuti panduan Jurusan Teknik Informatika Universitas Sriwijaya sebagai berikut.

BAB I. PENDAHULUAN

Bab ini memberikan penjelasan mengenai latar belakang, rumusan permasalahan yang akan dipecahkan, batasan-batasan dalam penelitian, serta tujuan dan manfaat dari penelitian yang berkaitan dengan sistem klasifikasi *hoax* menggunakan algoritma LSTM.

BAB II. KAJIAN LITERATUR

Pada bab ini akan disajikan tinjauan literatur yang terkait dengan topik deteksi berita *hoax* menggunakan metode LSTM. Kajian literatur ini mengacu pada beberapa penelitian sebelumnya.

BAB III. METODOLOGI PENELITIAN

Pada bab ini dijabarkan secara rinci metode yang digunakan untuk menjalankan penelitian ini. Penjelasan ini mencakup prosedur yang digunakan untuk mempersiapkan data berita *hoax*, menerapkan LSTM, dan seleksi model yang digunakan untuk mencapai tujuan penelitian ini.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini, akan dijelaskan tentang tahapan perancangan perangkat lunak yang akan diproduksi, dimulai dari analisis kebutuhan, perancangan, proses

konstruksi, dan akhirnya pengujian untuk memverifikasi kesesuaian perangkat lunak dengan kebutuhan penelitian.

BAB V. HASIL DAN PEMBAHASAN

Pada bab ini, akan disajikan penjelasan mengenai hasil yang telah diperoleh pada langkah sebelumnya. Data yang telah diuji akan dianalisis menggunakan beragam metode, dan hasilnya akan divalidasi.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini terdapat rangkuman dan saran-saran yang disusun oleh penulis sebagai hasil dari jawaban atas setiap tujuan dan manfaat dari penelitian ini.

1.8 Kesimpulan

Pada bab ini telah dijelaskan rencana penelitian, yang mencakup latar belakang penelitian, rumusan masalah, tujuan dan manfaat dari penelitian, batasan masalah, dan sistematika penulisan yang dibuat.

DAFTAR PUSTAKA

- Arief, B., Kholifatullah, H., & Prihanto, A. 2023. Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech. *Journal of Informatics and Computer Science*, 04.
- Bahad, P., Saxena, P., & Kamal, R. 2019. Fake News Detection using Bi-directional LSTM-Recurrent Neural Network. *Procedia Computer Science*, 165, 74–82. <https://doi.org/10.1016/j.procs.2020.01.072>
- Bandyopadhyay, S., Sharma, D. S., & Sangal, R. 2017. An Exploration of Word Embedding Initialization in Deep-Learning Tasks. In *NLP Association of India. NLP AI*. <https://code.google.com/archive/p/>
- Bengio, Y., Simard, P., & Frasconi, P. 1994. Learning Long-Term Dependencies with Gradient Descent is Difficult. In *IEEE Transactions On Neural Networks* (Vol. 5, Issue 2).
- Chen, C.-W., Tseng, S.-P., & Wang, J.-F. 2021. Outpatient Text Classification System Using LSTM. *Journal Of Information Science And Engineering*, 37, 365–379. <https://doi.org/10.6688/JISE.202103>
- Fauzi, A., Setiawan, E. B., & Baizal, Z. K. A. 2019. Hoax News Detection on Twitter using Term Frequency Inverse Document Frequency and Support Vector Machine Method. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012025>
- Jang, B., Kim, I., & Kim, J. W. 2019. Word2vec convolutional neural networks for classification of news articles and tweets. *Plos One*, 14(8). <https://doi.org/10.1371/journal.pone.0220976>
- Kumar Sharma, S., & Kishor Sharma, N. 2019. Text Classification using LSTM based Deep Neural Network Architecture. *International Journal on Emerging Technologies*, 10(4), 38–42. www.researchtrend.net
- Mah, P. M., Skalna, I., & Muzam, J. 2022. Natural Language Processing and Artificial Intelligence for Enterprise Management in the Era of Industry 4.0. *Applied Sciences (Switzerland)*, 12(18). <https://doi.org/10.3390/app12189207>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. 2013. Efficient Estimation of Word Representations in Vector Space. <http://arxiv.org/abs/1301.3781>
- Nayoga, B. P., Adipradana, R., Suryadi, R., & Suhartono, D. 2021. Hoax Analyzer for Indonesian News Using Deep Learning Models. *Procedia Computer Science*, 179, 704–712. <https://doi.org/10.1016/j.procs.2021.01.059>

- Orosoo, M., Govindasamy, S., Bayarsaikhan, N., Rajkumari, Y., Fatma, G., Manikandan, R., & Kiran Bala, B. 2023. Performance analysis of a novel hybrid deep learning approach in classification of quality-related English text. *Measurement: Sensors*, 28. <https://doi.org/10.1016/j.measen.2023.100852>
- Pramesti, D. D., Novitasari, D. C. R., Setiawan, F., & Khaulasari, H. 2022. Long-Short Term Memory (LSTM) For Predicting Velocity And Direction Sea Surface Current On Bali Strait. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 16(2), 451–462. <https://doi.org/10.30598/barekengvol16iss2pp451-462>
- Sari, W. K., Rini, D. P., & Firsandaya Malik, R. 2020. Multilabel Classification for News Article Using Long Short-Term Memory. In *Sriwijaya Journal of Informatic and Applications* (Vol. 01, Issue 01). <http://sjia.ejournal.unsri.ac.id>
- Srivastava, N., Mansimov, E., & Salakhutdinov, R. 2015. Unsupervised Learning of Video Representations using LSTMs. <http://arxiv.org/abs/1502.04681>
- Wang, J.-H., Liu, T.-W., Luo, X., & Wang, L. 2018. An LSTM Approach to Short Text Sentiment Classification with Word Embeddings.