

**KLASIFIKASI TEKS BERITA NON-FORMAL DAN FORMAL
DALAM BAHASA INDONESIA MENGGUNAKAN
ALGORITMA LONG SHORT-TERM MEMORY
BERDASARKAN KAMUS BESAR BAHASA INDONESIA**

Diajukan Sebagai Syarat Untuk Menyelesaikan

Pendidikan Program Strata-1 Pada

Jurusan Teknik Informatika



Oleh :

Diah Shinta Resmy

NIM : 09021282025089

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**

2024

LEMBAR PENGESAHAN SKRIPSI

KLASIFIKASI TEKS BERITA NON-FORMAL DAN FORMAL DALAM BAHASA INDONESIA MENGGUNAKAN ALGORITMA LONG SHORT-TERM MEMORY BERDASARKAN KAMUS BESAR BAHASA INDONESIA

Oleh :

Diah Shinta Resmy

NIM : 09021282025089

Palembang, 8 November 2024

Pembimbing 1



Alvi Syahrini Utami, S.Si., M.Kom.
NIP. 197812222006042003

Pembimbing 2



Desty Rodiah, S.Kom., M.T.
NIP. 19812212020122011

Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.

NIP. 198004182020121001

TANDA LULUS UJIAN KOMPREHENSIF

Pada hari Jum'at tanggal 04 Oktober 2024 Telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Diah Shinta Resmy

NIM : 09021282025089

Judul : Klasifikasi Teks Berita Non-Formal dan Formal dalam Bahasa Indonesia
Menggunakan Algoritma *Long Short-Term Memory* Berdasarkan Kamus Besar Bahasa Indonesia

Dan dinyatakan LULUS

1. Ketua Penguji

Yunita, M.Cs.

NIP. 198306062015042002

2. Penguji I

Novi Yusliani, M.T.

NIP. 198211082012122001

3. Pembimbing I

Alvi Syahrini Utami, S.Si., M.Kom.

NIP. 197812222006042003

4. Pembimbing II

Desty Rodiah, S.Kom., M.T.

NIP. 19812212020122011

.....
.....
.....
.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.

NIP. 198004182020121001

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini:

Nama : Diah Shinta Resmy

NIM : 09021282025089

Judul : Klasifikasi Teks Berita Non-Formal dan Formal dalam Bahasa Indonesia

Menggunakan Algoritma *Long Short-Term Memory* Berdasarkan Kamus

Besar Bahasa Indonesia

Hasil Pengecekan *Software* iThenticate/Turnitin : 16%

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya akan bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

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



Palembang, 8 November 2024



Diah Shinta Resmy

NIM. 09021282025089

MOTTO DAN PERSEMBAHAN

Motto:

إِنَّ مَعَ الْعُسْرِ يُسْرًا

“Sesungguhnya beserta kesulitan ada kemudahan.”

- Al-Insyirah (5): 6

Kupersembahkan Karya Tulis ini kepada:

- Allah SWT
- Orang Tua
- Keluarga Besar
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

This research aims to develop an Indonesian news text classification system to distinguish between formal and non-formal texts. This classification system utilises the Long Short-Term Memory (LSTM) algorithm with reference to the Big Indonesian Dictionary (KBBI). The development process involves text preprocessing techniques, such as data cleaning, lower case, remove punctuation, tokenising, stopword removal and stemming, to improve the quality of the data before it is processed by the LSTM model. The data used consists of 50,000 news texts divided into three parts: 60% training data, 20% validation data, and 20% testing data. The LSTM model was developed with hyperparameters such as RMSProp optimiser, batch size 128, and 40 epochs. The evaluation showed that the LSTM model was able to achieve 81.12% accuracy, 83.65% precision, 92.15% recall, and 87.70% F1-Score. However, the main challenges in this classification include the frequently changing variation of non-formal language, mislabelling in the dataset, limited representative data, risk of overfitting, as well as optimal hyperparameter determination, this system managed to overcome these challenges well.

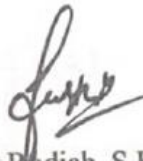
Keywords : Classification, NLP, KBBI, LSTM.

Supervisor 1



Alvi Syahrini Utami, S.Si., M.Kom.
NIP. 197812222006042003

Supervisor 2



Desty Rodiah, S.Kom., M.T.
NIP. 19812212020122011

Approve,

Head of Informatics Engineering Department



Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

ABSTRAK

Penelitian ini bertujuan mengembangkan sistem klasifikasi teks berita berbahasa Indonesia untuk membedakan antara teks formal dan non-formal. Sistem klasifikasi ini memanfaatkan algoritma *Long Short-Term Memory* (LSTM) dengan acuan Kamus Besar Bahasa Indonesia (KBBI). Proses pengembangan melibatkan teknik *preprocessing* teks, seperti *data cleaning*, *lower case*, *remove punctuation*, *tokenizing*, *stopword removal* dan *stemming*, untuk meningkatkan kualitas data sebelum diproses oleh model LSTM. Data yang digunakan terdiri dari 50.000 teks berita yang dibagi menjadi tiga bagian: 60% data *training*, 20% data *validation*, dan 20% data *testing*. Model LSTM dikembangkan dengan hyperparameter seperti *optimizer* RMSProp, *batch size* 128, dan 40 *epochs*. Evaluasi menunjukkan bahwa model LSTM mampu mencapai akurasi 81.12%, *precision* 83.65%, *recall* 92.15%, dan F1-Score 87.70%. Namun, tantangan utama dalam klasifikasi ini termasuk variasi bahasa non-formal yang sering berubah, kesalahan pelabelan dalam dataset, keterbatasan data yang representatif, resiko *overfitting*, serta penentuan *hyperparameter* yang optimal, sistem ini berhasil mengatasi tantangan tersebut dengan baik.

Kata Kunci : Klasifikasi, NLP, KBBI, LSTM.

Pembimbing 1



Alvi Syahrini Utami, S.Si., M.Kom.
NIP. 197812222006042003

Pembimbing 2



Desty Rodiah, S.Kom., M.T.
NIP. 19812212020122011

Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.

NIP. 198004182020121001

KATA PENGANTAR

Puji syukur kepada Allah SWT atas nikmat, kesehatan, kesempatan, dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian berjudul “Klasifikasi Teks Berita Non-Formal dan Formal dalam Bahasa Indonesia menggunakan Algoritma *Long Short-Term Memory* Berdasarkan Kamus Besar Bahasa Indonesia” sebagai syarat kelulusan Strata-1 (S1) di Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis juga mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan, bimbingan, dan bantuan, baik secara langsung maupun tidak langsung, dalam penyusunan skripsi ini.

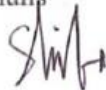
1. Allah SWT atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik.
2. Muhammad Rusli, S.Kom., M.Kom., sebagai ayah dan Ir. Yunia Ulfa sebagai ibunda tercinta atas segala doa, cinta, dukungan, serta pengorbanan yang tanpa henti. Semangat dan bimbingan yang diberikan telah menjadi sumber kekuatan bagi penulis dalam menyelesaikan skripsi.
3. Diri sendiri, yang telah berjuang melewati berbagai badai dan tetap tegar serta tidak menyerah dalam situasi tersulit adalah bukti dari kekuatan dan keteguhan yang telah membawa penulis sampai ke tahap ini.
4. Saudara penulis, Muhammad Satari Zaini atas pemberian semangat dan dukungan kepada penulis.

5. Bapak Prof. Dr. Erwin, S.Si., M.Si., sebagai Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Hadipurnawan Satria, Ph.D., sebagai Ketua Jurusan Teknik Informatika di Universitas Sriwijaya.
7. Ibu Alvi Syahrini Utami, S.Si., M.Kom., dan Ibu Desty Rodiah, S.Kom., M.T., selaku Dosen Pembimbing 1 dan 2, atas bimbingan serta kesabarannya dalam membantu penulis menyelesaikan skripsi.
8. Teman-teman penulis, Riska Trilia Utami, Atikah Syifa Warastri, Munawaroh Syahfitri, Afifah Putri Mevtalina dan Septry Lyman Thasya, serta teman-teman lainnya yang tidak bisa disebutkan satu-persatu, telah memberikan semangat dan kekuatan saat menghadapi berbagai tantangan selama penyusunan skripsi.
9. Seluruh staf Jurusan Teknik Informatika dan Fakultas Ilmu Komputer Sriwijaya.

Penulis menyadari skripsi ini masih memiliki kekurangan karena keterbatasan ilmu dan pengalaman. Oleh sebab itu, penulis mengharapkan kritik dan saran yang membangun untuk perbaikan ke depan. Semoga skripsi ini bermanfaat bagi semua pihak. Terima kasih.

Palembang, 8 November 2024

Penulis



Diah Shinta Resmy

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
TANDA LULUS UJIAN KOMPREHENSIF.....	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT.....	vi
ABSTRAK.....	vii
KATA PENGANTAR	viii
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I.....	I-1
PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-5
1.4 Tujuan Penelitian.....	I-5
1.5 Manfaat Penelitian.....	I-6
1.6 Batasan Masalah.....	I-6
1.7 Sistematika Penulisan	I-6
1.8 Kesimpulan.....	I-7
BAB II.....	II-1
KAJIAN LITERASI	II-1
2.1 Pendahuluan	II-1
2.2 Landasan Teori	II-1
2.2.1 Natural Language Processing	II-1
2.2.2 Klasifikasi Teks.....	II-2
2.2.3 Preprocessing.....	II-2
2.2.4 Word2vec	II-6
2.2.5 Skip-Gram.....	II-6
2.2.6 Long-Short Term Memory	II-12
2.2.7 Bahasa Non-Formal dan Formal dalam Bahasa Indonesia	II-15
2.2.8 Confusion Matrix	II-16

2.2.9	Rational Unified Process	II-18
2.3	Penelitian Lain yang Relevan.....	II-21
2.3.1	Implementasi Model Long-Short Term Memory (LSTM) pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia (Pratama, 2022).....	II-21
2.3.2	Improving the Accuracy of Text Classification using Stemming Method, a Case of Non-Formal Indonesian Conversation (Rianto et al. 2021)	II-22
2.3.3	Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan (Fajri and Syaiful 2022).....	II-23
2.4	Kesimpulan	II-23
BAB III.....		III-1
METODOLOGI PENELITIAN.....		III-1
3.1	Pendahuluan	III-1
3.2	Pengumpulan Data.....	III-1
3.2.1	Jenis Data.....	III-1
3.2.2	Sumber Data	III-1
3.2.3	Metode Pengumpulan Data.....	III-2
3.3	Tahapan Penelitian	III-2
3.3.1	Kriteria Pengujian	III-8
3.3.2	Format Data Pengujian.....	III-9
3.3.3	Alat yang Digunakan dalam Pelaksanaan Penelitian	III-9
3.4	Metode Pengembangan Perangkat Lunak	III-10
3.4.1	Fase Insepsi.....	III-10
3.4.2	Fase Elaborasi.....	III-11
3.4.3	Fase Konstruksi.....	III-11
3.4.4	Fase Transisi	III-11
3.5	Manajemen Proyek Penelitian	III-12
3.5.1	Penjadwalan Waktu Penelitian.....	III-12
3.6	Kesimpulan.....	III-19
BAB IV.....		IV-1
PENGEMBANGAN PERANGKAT LUNAK.....		IV-1
4.1	Pendahuluan	IV-1
4.2	<i>Rational Unified Process (RUP)</i>	IV-1
4.2.1	Fase Insepsi.....	IV-1
4.2.2	Fase Elaborasi.....	IV-10
4.2.3	Fase Konstruksi.....	IV-16

4.2.4	Fase Transisi	IV-22
4.3	Kesimpulan.....	IV-28
BAB V.....		V-1
HASIL DAN ANALISIS PENELITIAN		V-1
5.1	Pendahuluan	V-1
5.2	Data Hasil Penelitian	V-1
5.2.1	Data Konfigurasi Pengujian.....	V-1
5.2.2	Data Hasil Konfigurasi Pengujian.....	V-10
5.3	Analisis Hasil Penelitian.....	V-12
5.3.1	Hasil Pengujian Algoritma LSTM	V-12
5.4	Kesimpulan.....	V-18
BAB VI.....		VI-1
KESIMPULAN DAN SARAN.....		VI-1
6.1	Kesimpulan.....	VI-1
6.2	Saran.....	VI-1
DAFTAR PUSTAKA		xiv

DAFTAR TABEL

Tabel II-1. <i>Lower Case</i>	II-ii
Tabel II-2. <i>Remove Punctuation</i>	II-ii
Tabel II-3. <i>Tokenizing</i>	II-4
Tabel II-4. <i>Stemming</i>	II-5
Tabel II-5. Bentukan Kata.....	II-16
Tabel II-6. <i>Confusion Matrix</i>	II-17
Tabel III-1. Rancangan Tabel Percobaan <i>Hyperparameter</i>	III-8
Tabel III-2. Rancangan Tabel Metrik Evaluasi.....	III-8
Tabel III-3. <i>Work Breakdown Structure (WBS)</i>	III-13
Tabel IV-1. Kebutuhan Fungsional.....	IV-3
Tabel IV-2. Kebutuhan Non Fungsional.....	IV-3
Tabel IV-3. Definisi <i>User Use Case</i>	IV-6
Tabel IV-4. Definisi Diagram <i>Use Case</i>	IV-6
Tabel IV-5. Skenario Mengklasifikasi dengan Algoritma LSTM.....	IV-7
Tabel IV-6. Skenario Visualisasi Data.....	IV-8
Tabel IV-7. Implementasi Kelas.....	IV-21
Tabel IV-8. Rencana Pengujian <i>Use Case</i> Mengklasifikasi dengan Algoritma LSTM.....	IV-27
Tabel IV-10. Pengujian <i>Use Case</i> Visualisasi Data.....	IV-28
Tabel V-1. Hasil Analisis Pengujian dari <i>Data Training</i>	V-2
Tabel V-2. Hasil Analisis Pengujian dari <i>Data Testing</i>	V-5
Tabel V-3. <i>Confusion Matrix</i>	V-10
Tabel V-4. Hasil Rancangan Tabel Percobaan <i>Hyperparameter</i>	V-12
Tabel V-5. Hasil <i>Accuracy Data Training</i>	V-16
Tabel V-6. Hasil Rancangan Tabel Metrik Evaluasi.....	V-16

DAFTAR GAMBAR

Gambar II- 1. <i>Skip-Gram</i>	II-7
Gambar II- 2. Ilustrasi <i>Window Model Skip-Gram</i>	II-8
Gambar II- 3. Arsitektur LSTM.....	II-12
Gambar II- 4. <i>Memory Cell LSTM</i>	II-13
Gambar II- 5. <i>Rational Unified Process</i>	II-19
Gambar III- 1. Tahapan Penelitian.....	III-2
Gambar III- 2. Diagram Alir Sistem.....	III-4
Gambar IV- 1. Diagram <i>Use Case</i>	IV-5
Gambar IV- 2. Rancangan <i>Interface Dashboard</i>	IV-11
Gambar IV- 3. Rancangan <i>Interface Menu Data</i>	IV-12
Gambar IV- 4. Rancangan <i>Interface Hasil Diklasifikasikan</i>	IV-12
Gambar IV- 5. Diagram <i>Sequence</i> Mengklasifikasi dengan Algoritma LSTM.....	IV-14
Gambar IV- 6. Diagram <i>Sequence</i> Menampilkan Visualisasi Data.....	IV-15
Gambar IV- 7. Mockup & Layout GUI.....	IV-17
Gambar IV- 8. Diagram Kelas.....	IV-19
Gambar IV- 9. Implementasi <i>Interface Dashboard</i>	IV-19
Gambar IV- 10. Implementasi <i>Interface Menu Data</i>	IV-20
Gambar IV- 11. Implementasi <i>Interface Hasil Diklasifikasikan</i>	IV-20

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab pendahuluan ini akan membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan, serta kesimpulan dalam kegiatan penelitian. Bab ini akan menjelaskan penjelasan umum mengenai keseluruhan penelitian dengan menggunakan algoritma *Long Short-Term Memory* (LSTM) dalam kasus klasifikasi teks berita non-formal dan formal berdasarkan Kamus Besar Bahasa Indonesia (KBBI). Pendahuluan dimulai dengan penjelasan masalah yang ada dan bagaimana penyelesaian yang akan dicoba.

1.2 Latar Belakang Masalah

Di era globalisasi ini, terdapat beberapa pengaruh yang mengubah keaslian bahasa Indonesia. Era globalisasi adalah masa perubahan global yang melanda seluruh dunia, termasuk Indonesia. Dampaknya sangat besar terhadap berbagai aspek kehidupan, termasuk bahasa. Perubahan dalam aspek bahasa dipicu oleh kemajuan teknologi. Teknologi yang semakin canggih di era globalisasi memengaruhi berbagai aspek. Salah satu alasan kemajuan teknologi dapat memicu perubahan dalam aspek bahasa adalah karena teknologi memudahkan masuknya bahasa asing, sehingga bahasa asing lebih mudah menyusup ke Indonesia (Desi, 2022). Salah satunya adalah media massa daring Hipwee.com yang menggunakan bahasa non-formal. Tanggung jawab media dalam menyampaikan informasi secara

tepat dan baik dapat mempengaruhi perilaku masyarakat Indonesia secara luas dan mendalam. Oleh karena itu, penelitian terkait perilaku media, termasuk bahasa yang digunakan, menjadi sangat penting (Dewi, 2019). Pentingnya penggunaan bahasa adalah untuk memastikan informasi yang disampaikan mudah dipahami dan tidak disalahartikan oleh masyarakat. Selain itu, informasi yang disajikan dalam bentuk berita juga harus memiliki nilai serta memberikan manfaat (Maijulika *et al.*, 2023).

Bahasa formal ditandai dengan penggunaan kata-kata yang tepat, struktur kalimat yang tersusun rapi, serta penghindaran penggunaan slang atau bahasa sehari-hari (Vebrina *et al.*, 2024). Bahasa formal umumnya digunakan dalam konteks seperti penulisan jurnal, laporan penelitian, dan sejenisnya. Sementara itu, Bahasa non-formal digunakan dalam percakapan santai dengan teman, *chatting*, berkomentar, atau membuat postingan di media sosial, dan lain sebagainya (Qulub *et al.*, 2019).

Proses pengklasifikasian merujuk pada kegiatan mempelajari dan menganalisis dokumen teks yang telah diklasifikasikan sebelumnya, dengan tujuan menghasilkan model yang dapat digunakan untuk mengelompokkan dokumen teks baru yang belum memiliki kategori. Klasifikasi adalah proses yang digunakan untuk mengelompokkan dokumen ke dalam kategori atau kelas yang sesuai berdasarkan karakteristik dokumen tersebut. Umumnya, pelabelan dalam proses klasifikasi masih dilakukan secara manual oleh tim ahli, terutama pada dataset yang berukuran besar (Habib *et al.*, 2022). Oleh karena itu, penelitian ini bertujuan untuk mengembangkan proses pengklasifikasian yang dapat melakukan pelabelan otomatis dalam skala besar sebagai pengganti pelabelan manual.

Hal ini menimbulkan kebutuhan untuk mengembangkan sistem otomatis yang dapat mengklasifikasikan berita secara efisien berdasarkan kategori formal dan non-formal. Pendekatan ini dipilih untuk mengembangkan metode klasifikasi teks berita formal dan non-formal dalam Bahasa Indonesia berbasis *Natural Language Processing* (NLP) yang mengacu pada Kamus Besar Bahasa Indonesia (KBBI). Melalui pendekatan ini, diharapkan dapat dihasilkan sebuah sistem yang mampu mengidentifikasi kategori teks secara otomatis dengan tingkat akurasi yang tinggi. Implementasi NLP dalam penelitian ini akan mencakup analisis dan pemrosesan teks untuk membedakan antara bahasa formal dan non-formal, serta memberikan hasil klasifikasi yang sesuai berdasarkan aturan bahasa yang tercantum dalam KBBI.

Long Short-Term Memory (LSTM) merupakan metode yang lebih baik dibandingkan metode tradisional, sangat cocok untuk analisis sentimen dan klasifikasi teks. Oleh karena itu, algoritma LSTM sebagai metode yang tepat karena keunggulannya dalam menangani data temporal dan urutan (Daiman *et al.* 2024). Adapun metode LSTM kelebihanannya adalah kemampuan dalam melakukan klasifikasi teks dengan lebih baik karena dapat memproses interaksi antar teks secara efektif. Selain itu, proses pada LSTM mampu menentukan nilai yang akan dipilih sebagai keluaran sehingga menjadi masukan yang relevan (Fajri dan Syaiful, 2022).

Dalam beberapa penelitian dengan metode yang sama pada peneliti terdahulu. Penelitian pertama oleh (Ikhsan, 2023) penelitian ini menganalisis sentimen pengguna Twitter terhadap kenaikan harga bahan bakar minyak pada bulan Juli

2022 menggunakan model *Long Short-Term Memory* (LSTM). Data dikumpulkan dari Twitter melalui *scraping* dengan kata kunci terkait, lalu diproses dan diberi label untuk membedakan sentimen positif dan negatif. Model LSTM, yang mampu menangani data sekuensial dengan baik, digunakan dalam lima percobaan untuk memprediksi sentimen. Hasil evaluasi menunjukkan bahwa model ini memiliki akurasi bervariasi antara 86%-90%, dengan *precision* dan *recall* yang tinggi untuk kedua jenis sentimen. Model ini menunjukkan bahwa sentimen positif cenderung lebih dominan dibandingkan sentimen negatif dalam sebagian besar percobaan. Analisis sentimen ini berguna bagi perusahaan dan organisasi untuk memahami opini publik secara cepat dan efektif, serta menunjukkan bahwa model *deep learning* seperti LSTM lebih unggul dibandingkan model tradisional dalam analisis sentimen di media sosial. Penelitian selanjutnya, oleh (Nabbani *et al.*, 2020) penelitian ini bertujuan untuk menganalisis sentimen data *tweet* Provider X sebanyak 5.375 dari 24 Desember 2019 - 24 Februari 2020. Data yang dikumpulkan diklasifikasikan ke dalam dua kelas, negatif dan positif. Penelitian ini menggunakan metode *Long Short-Term Memory*. Pengujian *hyperparameter* dilakukan untuk mendapatkan model yang paling optimal. Hasil pengujian menghasilkan *hyperparameter* Word2vec termasuk dimensi 50, *window size* 5, *epoch* 50, dan *learning rate* 0.025. Pada *hyperparameter* LSTM, kondisi optimal diperoleh ketika jumlah lapisan LSTM adalah satu, lima unit LSTM, fungsi aktivasi adalah sigmoid, optimizer adalah Nadam, ukuran *dropout* rekuren adalah 0,2, dan ukuran *batch* adalah 32. Model LSTM dievaluasi menggunakan K-fold cross-validation dan menghasilkan akurasi rata-rata sebesar 97,53%, *f1-score* 97,39%, dan *loss* 10,36%.

Berdasarkan hasil dari penelitian yang telah dijelaskan sebelumnya, dapat disimpulkan bahwa nilai akurasi dari pemodelan algoritma LSTM sudah melebihi 50% sehingga bisa dijadikan referensi untuk melakukan penelitian dengan menggunakan model *Long Short-Term Memory* (LSTM). Maka dari itu, penelitian tersebut akan menjadi landasan pada penelitian ini. Suatu harapan bahwa penelitian ini dapat lebih memperkuat klasifikasi teks berita non-formal dan formal dalam bahasa Indonesia berdasarkan Kamus Besar Bahasa Indonesia bahwa penggunaan algoritma LSTM menunjukkan tingkat akurasi yang baik untuk digunakan dalam klasifikasi.

1.3 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah dalam penelitian ini yaitu sebagai berikut:

1. Bagaimana membangun sistem untuk teks non-formal dan formal dalam klasifikasi dalam algoritma *Long Short-Term Memory*?
2. Bagaimana kinerja algoritma *Long Short-Term Memory* dalam mengklasifikasikan teks non-formal dan formal?

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini berdasarkan rumusan masalah yaitu sebagai berikut :

1. Menghasilkan sistem klasifikasi yang efektif dan efisien dengan model LSTM dengan arsitektur dan *hyperparameter* pada teks non-formal dan formal.

2. Mengetahui kinerja algoritma *Long Short-Term Memory* (LSTM) yang dinilai melalui *accuracy*, *precision*, *recall*, dan *f1-score*.

1.5 Manfaat Penelitian

Pada penelitian ini, dapat memberikan manfaat berupa :

1. Bagi penulis, dapat mengeksplorasi bahasa dalam pengenalan teks berita non-formal dan formal, sehingga dapat menambah pengetahuan dan pemahaman tentang bahasa yang diterapkan berbagai bidang.
2. Bagi Pengembang lain, dapat memanfaatkan hasil penelitian ini untuk membangun sistem klasifikasi otomatis yang mampu memisahkan berita formal dan non-formal dengan lebih akurat, meningkatkan efisiensi moderasi atau kurasi konten.

1.6 Batasan Masalah

Batasan masalah yang ditetapkan pada penelitian ini, adalah :

1. Metode yang digunakan untuk pembobotan teks berita non-formal dan formal adalah *Word2vec Skip-Gram*.
2. Data yang digunakan dalam penelitian berasal dari platform daring *Kaggle.com* dalam bahasa Indonesia yang sudah ada.

1.7 Sistematika Penulisan

Dalam penulisan penelitian ini, penulis membuat sistematika penulisan standar Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang terdiri dari 3 bab, dengan uraian masing-masing bab sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Bab ini berisi tentang kajian pustaka atau referensi terkait topik penelitian yang dilakukan, dengan tujuan untuk mendukung argumentasi dan memberikan landasan teori yang kuat.

BAB III. METODOLOGI PENELITIAN

Bab ini berisi tentang rancangan penelitian yang dilakukan, termasuk jenis penelitian, populasi dan sampel penelitian, teknik pengumpulan data, analisis data, dan tahapan penelitian.

1.8 Kesimpulan

Pada penelitian ini, berfokus pada pengembangan sistem untuk mengklasifikasikan teks berita non-formal dan formal dalam bahasa Indonesia menggunakan algoritma *Long Short-Term Memory* (LSTM) dengan acuan Kamus Besar Bahasa Indonesia (KBBI). Masalah utama yang dihadapi adalah sulitnya membedakan bahasa formal dan non-formal dalam berita *online* dan kebutuhan akan sistem otomatis untuk klasifikasi berita. Tujuan dari penelitian ini adalah untuk membangun sistem klasifikasi yang efektif serta mengevaluasi kinerja algoritma LSTM dalam tugas tersebut. Penelitian ini diharapkan dapat memberikan manfaat, seperti meningkatkan pemahaman masyarakat tentang bahasa formal dan non-formal, membantu peneliti lain dalam penulisan berita formal yang baik.

DAFTAR PUSTAKA

- AGUSTINA, REGITA, RITA MAGDALENA, and NOR KUMALASARI CAECAR PRATIWI. 2022. "Klasifikasi Kanker Kulit Menggunakan Metode Convolutional Neural Network Dengan Arsitektur VGG-16." *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika* 10(2): 446. doi:10.26760/elkomika.v10i2.446.
- Asy Syifa, Siti, and Irma Amelia Dewi. 2022. "MIND (Multimedia Artificial Intelligent Networking Database Arsitektur Resnet-152 Dengan Perbandingan Optimizer Adam Dan RMSProp Untuk Mendeteksi Penyakit Paru-Paru." *Journal MIND Journal | ISSN* 7(2): 139–50. <https://doi.org/10.26760/mindjournal.v7i2.139-150>.
- Budiman, Ariel Elbert, and Andreas Widjaja. 2020. "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir." *Jurnal Teknik Informatika dan Sistem Informasi* 6(3): 475–88. doi:10.28932/jutisi.v6i3.2892.
- Chandra, Albert Yakobus, Didik Kurniawan, and Rahmat Musa. 2020. "Perancangan Chatbot Menggunakan Dialogflow Natural Language Processing (Studi Kasus: Sistem Pemesanan Pada Coffee Shop)." *Jurnal Media Informatika Budidarma* 4(1): 208. doi:10.30865/mib.v4i1.1505.
- Daiman, Claudia Nila, Aviv Yuniar Rahman, Firman Nudiyansyah, Program Studi, Teknik Informatika, Fakultas Teknik, Universitas Widyagama Malang, et al. 2024. "KLASIFIKASI TEKS BERITA BREAKING NEWS DI MANGGARAI." 7(2): 170–74.
- Darmawan, Doni. 2023. *Implementasi Metode Convolutional Neural Network (CNN) Dalam Mendeteksi Jenis Sampah*.
- Desi, Karolina. 2022. "Dampak Perkembangan Bahasa Asing Terhadap Bahasa Indonesia Di Era Globalisasi." *Jurnal Pendidikan Tambusai* 6(1): 2569–77. <https://www.jptam.org/index.php/jptam/article/view/3312>.
- Dewi, Mariani. 2014. "Gaya Bahasa Berita Media Online Di Indonesia: Judul Menarik Tidak Harus Tidak Baku." *Humaniora* 5(2): 1015. doi:10.21512/humaniora.v5i2.3212.
- Dwi Anggita, Kenza, Yuita Arum Sari, and Sigit Adinugroho. 2021. "Analisis Sentimen Terhadap Opini Masyarakat Mengenai Kebijakan PSBB Menggunakan Metode Naive Bayes Dengan Seleksi Fitur Improved Gini Index." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 5(12): 5288–93. <http://j-ptiik.ub.ac.id>.
- Dwi Pratama, Erico. 2022. "Implementasi Model Long-Short Term Memory (LSTM) Pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia." *The Journal on Machine Learning and Computational Intelligence (JMLCI)* 1(2): 38–42.

- Fajri, Fathorazi Nur, and Syaiful Syaiful. 2022. "Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan." *Building of Informatics, Technology and Science (BITS)* 4(3): 1625–33. doi:10.47065/bits.v4i3.2635.
- Fauzan Ziqroh, Muhammad, and Edy Santoso. 2023. "Klasifikasi Jenis Barang Bekas Menggunakan Metode Naïve Bayes Dengan Seleksi Fitur Information Gain (Studi Kasus: Akun Instagram Jual Beli Barang Bekas @infobarkas_Jogja)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 7(1): 140–47. <http://j-ptiik.ub.ac.id>.
- Fitri Niasita, Ananda, Putra Pandu Adikara, and Sigit Adinugroho. 2019. "Analisis Sentimen Pembangunan Infrastruktur Di Indonesia Dengan Automated Lexicon Word2Vec Dan Naive-Bayes." *Jurnal Pengembangan Teknologi Informarsi dan Ilmu Komputer* 3(3): 2673–79. <http://j-ptiik.ub.ac.id>.
- Habib, Sayyid Muhammad, Elin Haerani, Siska Kurnia Gusti, and Siti Ramadhani. 2022. "Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier." *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)* 5(2): 248–58. doi:10.32672/jnkti.v5i2.4191.
- Hakim, Zaenal, and Robby Rizky. 2018. "249340042." 6(2): 103–12.
- Hartawan, George Pri. 2017. "IMPLEMENTASI RATIONAL UNIFIED PROCESS DALAM SISTEM INFORMASI E-SEKOLAH (Studi Kasus SMA Negeri 1 Cibadak)." *Jurnal SANTIKA: Jurnal Ilmiah Sains dan Teknologi* 7(1): 563–71.
- Ikhsan, Muhammad. 2023. "Analisis Sentimen Terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Long Short-Term Memory." *The Indonesian Journal of Computer Science Research* 2(1): 31–41. doi:10.59095/ijcsr.v2i1.29.
- Irma Purnamasari, Ade, and Irfan Ali. 2024. "Analisis Sentimen Komentar Berita Detik.Com Menggunakan Algoritma Suport Vektor Machine (Svm)." *Jurnal Mahasiswa Teknik Informatika* 8(3): 3175–81.
- Maitri, Agra Laksmi, and Joko Sutopo. 2019. "Rancangan Bangun Chatbot Sebagai Pusat Informasi Lembaga Kursus Dan Pelatihan Menggunakan Pendekatan Natural Language Processing." *Eprints.Uty.Ac.Id:* 1–9. <http://eprints.uty.ac.id/>.
- Maysanjaya, I Md. Dendi. 2020. "Klasifikasi Pneumonia Pada Citra X-Rays Paru-Paru Dengan Convolutional Neural Network." *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* 9(2): 190–95. doi:10.22146/jnteti.v9i2.66.
- Musfiroh, Desi, Ulfa Khaira, Pradita Eko Prasetyo Utomo, and Tri Suratno. 2021. "Analisis Sentimen Terhadap Perkuliahan Daring Di Indonesia Dari Twitter Dataset Menggunakan InSet Lexicon." *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 1(1): 24–33.

doi:10.57152/malcom.v1i1.20.

- Mutawalli, Lalu, Mohammad Taufan Asri Zaen, and Wire Bagye. 2019. "KLASIFIKASI TEKS SOSIAL MEDIA TWITTER MENGGUNAKAN SUPPORT VECTOR MACHINE (Studi Kasus Penusukan Wiranto)." *Jurnal Informatika dan Rekayasa Elektronik* 2(2): 43. doi:10.36595/jire.v2i2.117.
- Normawati, Dwi, and Surya Allit Prayogi. 2021. "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter." *Jurnal Sains Komputer & Informatika (J-SAKTI)* 5(2): 697–711.
- Nugroho, Pulung Adi, Indah Fenriana, and Rudy Arijanto. 2020. "Implementasi Deep Learning Menggunakan Convolutional Neural Network (CNN) Pada Ekspresi Manusia." *Algor* 2(1): 12–21.
- Nurdin, Arliyanti, Bernadus Anggo Seno Aji, Anugrayani Bustamin, and Zaenal Abidin. 2020. "Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks." *Jurnal Tekno Kompak* 14(2): 74. doi:10.33365/jtk.v14i2.732.
- Octaviano, Dimas, Program Studi, Teknik Informatika, Fakultas Sains, D A N Teknologi, Universitas Islam, and Negeri Syarif. 2019. *Perbandingan Seleksi Fitur Term Frequency*.
- Oktavianto, Hardian, and Rahman Puji Handri. 2020. "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes." *INFORMAL: Informatics Journal* 4(3): 117. doi:10.19184/isj.v4i3.14170.
- Prasetyo, Windra. 2019. "Algoritma Stemming Teks Bahasa Massenrempulu Berbasis Aturan Tata Bahasa."
- Putra, Febrianda, Hafsa Fulaila Tahiyat, Raja Muhammad Ihsan, Rahmaddeni Rahmaddeni, and Lusiana Efrizoni. 2024. "Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing Untuk Penentuan Keterangan Berat Badan Manusia." *MALCOM: Indonesian Journal of Machine Learning and Computer Science* 4(1): 273–81. doi:10.57152/malcom.v4i1.1085.
- Qulub, Mudawil, Ema Utami, Andi Sunyoto, Kata Kunci, Kata Imbuan, Tidak Baku, and Jaro-Winkler Distance. 2019. "Stemming Kata Berimbuhan Tidak Baku Bahasa Indonesia Menggunakan Algoritma Jaro-Winkler Distance Stemming Non-Formal Affix Word Indonesian Language Using Jaro-Winkler Distance Algorithm." : 254–63.
- Ramanizar, Haiqal, Al Fajri, Roy Binsar Sinaga, Hasan Mubarak, Albet Dwi Pangestu, and Desta Sandya Prasvita. 2021. "Analisis Sentimen Pengguna Twitter Terhadap Konflik Antara Palestina Dan Israel Menggunakan Metode Naïve Bayesian Classification Dan Support Vector Machine." *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA) Jakarta-Indonesia* (September): 166–75.

- Rezaputra, N, and Y D Prabowo. 2020. “Alih Bentuk Kalimat Non-Formal Menjadi Kalimat Formal Menggunakan Pendekatan Machine Translation.” *KALBISIANA Jurnal Sains, Bisnis dan ...* 8(1): 379–90. <http://112.78.142.42/index.php/kalbisiana/article/view/275%0Ahttp://112.78.142.42/index.php/kalbisiana/article/download/275/217>.
- Rianto, Achmad Benny Mutiara, Eri Prasetyo Wibowo, and Paulus Insap Santosa. 2021. “Improving the Accuracy of Text Classification Using Stemming Method, a Case of Non-Formal Indonesian Conversation.” *Journal of Big Data* 8(1): 1–16. doi:10.1186/s40537-021-00413-1.
- Tia, Tuti, Ilyas Nuryasin, and Maskur Maskur. 2020. “Model Simulasi Rational Unified Process (RUP) Pada Pengembangan Perangkat Lunak.” *Jurnal Repositor* 2(4): 485–94. doi:10.22219/repositor.v2i4.390.
- Vebrina, Kezia Youan, Lintang Kinasih, Khoilila Muddalifa, Luh Putu Restiana Wulandari, Moh Harun Al Rosid, and Dewi Puspa Arum. 2024. “Strategi Komunikasi Efektif Dalam Bisnis Penerapan Kalimat Yang Tidak Ambigu Untuk Menghindari Gagal Paham.” *Journal of Management and Innovation Entrepreneurship (JMIE)* 1(3): 492–97. doi:10.59407/jmie.v1i3.609.
- Wiranda, Laras, and Mujiono Sadikin. 2019. “Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. Metiska Farma.” *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)* 8(3): 184–96.
- Witanto, Karlina Surya, Ngurah Agus Sanjaya ER, AAIN Eka Karyawati, I Gusti Agung Gede Arya Kadyanan, I Ketut Gede Suhartana, and Luh Gede Astuti. 2022. “Implementasi LSTM Pada Analisis Sentimen Review Film Menggunakan Adam Dan RMSprop Optimizer.” *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)* 10(4): 351. doi:10.24843/jlk.2022.v10.i04.p05.