

KLASIFIKASI TEKS BERITA BREAKING NEWS DI MANGGARAI MENGUNAKAN LONG SHORT TERM MEMORY (LSTM)

Claudia Nila Daiman¹, Aviv Yuniar Rahman², Firman Nudiyanasyah³

^{1,2,3} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Widyagama Malang
claudianiladaiman8@gmail.com

ABSTRAK

Berita sering kali menyebar dari berbagai sumber, termasuk media sosial dan situs web. Metode LSTM (*Long Short Term Memory*) yang lebih baik dalam mengolah data temporal dan sekuensial dapat mempercepat pengambilan keputusan terhadap berita terkini di Manggarai. Tujuan dari penelitian ini adalah untuk mengembangkan sistem klasifikasi teks berita terkini menggunakan LSTM dan membuat model yang dapat mengklasifikasikan berita dengan akurasi tinggi ke dalam empat kategori: ekonomi, kecelakaan, politik dan pariwisata. Penelitian ini menggunakan 4000 dataset yang masing-masing kategori terdiri dari 1000 unit data. Data tersebut dibagi menjadi beberapa variasi rasio data latih dan uji: 3600:400, 3200:800, 2400:1600 dan 1600:2400. Model LSTM menunjukkan performa terbaik dengan rasio 3600:400, presisi 88,75%, presisi 88,79%, recall 88,75%, dan skor F1 88,76%. Akurasi menunjukkan persentase prediksi yang benar, precision mengukur ketepatan prediksi positif, recall menghitung seberapa baik model menangkap semua contoh positif, dan F1-score merupakan rata-rata harmonis dari precision dan recall. Hasil tersebut menunjukkan bahwa model LSTM dapat mengklasifikasikan teks berita secara efisien dan akurat. Penelitian ini memvalidasi penerapan LSTM dalam klasifikasi teks berita untuk memberikan informasi penting dan cepat kepada masyarakat Manggarai.

Keyword: LSTM, Klasifikasi Teks, Mangarai, Breaking News, Akurasi Tinggi, Media sosial.

1. PENDAHULUAN

Berita secara umum adalah informasi yang mengalir dari banyak sumber, banyak definisinya, namun ada beberapa kesimpulan yang dapat diambil mengenai pengertian berita.[1]. Perkembangan teknologi pada zaman sekarang ini sangatlah mempengaruhi kehidupan manusia.

Manggarai merupakan wilayah Nusa Tenggara Timur (NTT) yang termasuk wilayah yang cukup luas. Begitu banyak kejadian maupun kegiatan yang dijadikan berita atau opini yang dipaparkan di website atau biasa di sebut berita online. Mangarai, sebagai lokasi kajian, mungkin memiliki kekhasan dan preferensi informasi tersendiri. Oleh karena itu, penting untuk mengembangkan metode klasifikasi yang mempertimbangkan konteks lokal untuk memenuhi kebutuhan informasi yang khusus di daerah ini.

Salah satu penerapan klasifikasi teks yang dapat ditemukan adalah pengklasifikasian teks berita berdasarkan kategori tertentu[2]. Klasifikasi teks berita breaking news memiliki relevansi dan urgensi yang tinggi dalam konteks informasi lokal, dengan pentingnya mempertimbangkan aspek temporal dalam klasifikasi tersebut[3]. LSTM adalah metode yang lebih baik dibandingkan metode tradisional, metode LSTM sangat cocok untuk analisis sentimen dan klasifikasi teks[4]. Oleh karena itu, peneliti memilih LSTM sebagai metode yang tepat karena keunggulannya dalam menangani data temporal dan urutan.

Dalam penelitian ini, peneliti lain juga telah memberikan wawasan penting dengan mengeksplorasi aplikasi metode LSTM dalam klasifikasi teks berita. LSTM telah berhasil

menyelesaikan berbagai permasalahan, seperti *handwriting recognition*, *speech recognition*, *handwriting generation*, dan *image captioning*[5]. Hasil klasifikasi menunjukkan perhitungan akurasi terbaik sebesar 79% pada data latih dengan nilai epoch 27 dan nilai akurasi pada data validasi sebesar 77%[6].

Para peneliti fokus mengklasifikasikan teks berita terkini di Manggarai, membaginya menjadi empat kategori utama: politik, ekonomi, kecelakaan, dan pariwisata. Penelitian ini menggunakan data sekunder berupa berita online dari berbagai sumber pada tahun 2019 hingga 2024. Tujuan utama penelitian ini adalah untuk meningkatkan akurasi klasifikasi dan mempercepat proses penyuntingan dengan menyajikan berita terkini kepada pembaca. Feedback dari perusahaan berita mengenai hasil klasifikasi yang diperoleh memberikan para peneliti wawasan berharga tentang cara mengoptimalkan model klasifikasi mereka.

2. TINJAUAN PUSTAKA

2.1. *Breaking News*

Berita merupakan sejumlah peristiwa yang terjadi di dunia, tetapi hanya sebagian kecil saja yang dilaporkan[7]. Breaking news merupakan sebuah berita yang dianggap sangat penting, mendesak, dan memiliki dampak signifikan pada masyarakat. jenis berita ini harus disampaikan segera kepada semua masyarakat karena relevansinya yang tinggi.

2.2. Konteks Lokal Manggarai

Konteks lokal dalam bertutur ketika penutur berada dalam situasi budaya masyarakat tertentu

dapat dikaitkan dengan makna pragmatik tuturannya[8]. Dalam hal penggunaan Bahasa dalam berkomunikasi harus didasari konteks yang jelas. Konteks lokal Manggarai memiliki keunikan yang mencakup bahasa, budaya, dan preferensi masyarakat. Dalam aspek bahasa, penting untuk memahami dialek dan kosakata lokal agar informasi dapat disampaikan dengan akurat.

2.3. Klasifikasi Teks

Klasifikasi teks adalah pengklasifikasian data ke dalam kelompok atau kategori yang telah ditentukan[9]. Dalam hal ini, klasifikasi teks menggunakan data pelatihan berlabel untuk menangani banyak kasus guna mengklasifikasikan data pengujian ke dalam kelompok yang telah ditentukan sebelumnya. Proses klasifikasi adalah proses memperoleh suatu fitur atau model yang menjelaskan atau memisahkan konsep/kelas informasi. Setelah suatu pola ditemukan, pola tersebut dapat digunakan untuk memprediksi kelas data baru[10].

2.4. Preprocessing Teks

Pre-prosesing teks bertujuan untuk untuk mengubah data teks menjadi bentuk yang lebih baik sehingga menghasilkan informasi teks dengan kualitas yang baik dan siap digunakan pada proses selanjutnya[11]. Struktur data yang baik dapat menyederhanakan proses algoritma secara efektif dan otomatis. *Preprocessing* merupakan sebuah proses yang sangat penting dalam klasifikasi data teks[12]. Oleh karena itu diperlukan proses perubahan bentuk menjadi data terstruktur agar dapat diolah dengan lebih efektif.

2.5. Long Short-Term Memory (LSTM)

LSTM memiliki struktur kompleks dan unit memori yang dapat dilepas yang memungkinkan penyimpanan dan akses data jangka panjang[13]. Dalam konstruksi LSTM, terdapat beberapa komponen kunci, termasuk gerbang input, gerbang lupa, gerbang keluaran, bobot, dan bias. Struktur gerbang tersebut mencakup input gate, forget gate, dan output gate[14].

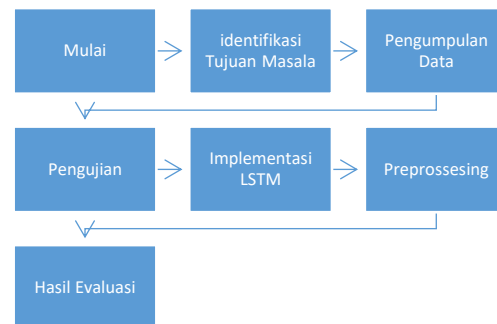
Pemilihan LSTM pada penelitian ini didasarkan pada keunggulan arsitektur tersebut dalam pengolahan data temporal dan sekuensial. LSTM memiliki struktur kompleks yang memungkinkan penyimpanan dan akses data jangka panjang, serta kemampuan untuk mengatasi masalah fuzzy. Hal ini membuat LSTM sangat cocok untuk menganalisis teks berita terkini dengan konteks temporal yang penting.

Dalam literatur, seperti yang dijelaskan dalam jurnal "Learning Stochastic Feedforward Neural Networks", LSTM telah menunjukkan hasil yang sangat baik dalam berbagai aplikasi, termasuk analisis teks dan pengenalan pola temporal. Tang dan Salakhutdinov berpendapat bahwa kemampuan

LSTM untuk memodelkan hubungan temporal yang kompleks menjadikannya pilihan yang tepat untuk banyak kasus penggunaan [15].

Penelitian ini menambah kontribusi baru terhadap penerapan LSTM dalam analisis teks berita terkini di wilayah Manggarai. Dengan menerapkan LSTM dalam konteks lokal, penelitian ini memberikan solusi baru untuk analisis teks berita yang efisien, dengan mempertimbangkan konteks temporal yang penting.

3. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

Gambar 1, diatas merupakan rancangan alur penelitian untuk klasifikasi teks berita breaking news di manggarai menggunakan algoritma LSTM. Dalam alur penelitian ini mencakup beberapa tahap, yang bertujuan untuk mencapai penelitian yang sistematis.

3.1 Identifikasi Tujuan Masalah

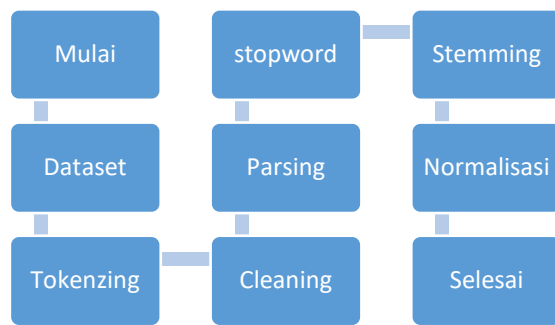
Identifikasih masalah merupakan proses menentukan tujuan yang di capai dalam menyelesaikan masalah. Dalam penelitian ini tujuannya adalah menghasilkan sebuah klasifikasi teks berita breaking news di manggarai berdasarkan empat kategori yaitu Kecelakaan, Politik, ekonomi, dan wisata.

3.2 Pengumpulan Data

Pengumpulan data penelitian ini menggunakan data sekunder yang datanya diambil dari internet yaitu teks berita terkini Manggarai yang diposting di Twitter, Instagram, Kompas.com, detik.com, Pemkab Manggarai, Suara. ya, Antanews. com dan media sosial lainnya terkait berita Manggarai terkini.

3.1. Preprocessing Data

Pemrosesan awal data adalah serangkaian operasi yang dilakukan untuk membersihkan, mengatur, dan menyiapkan data mentah sebelum diproses lebih lanjut atau diterapkan ke model pembelajaran mesin.



Gambar 2. Tahap Preprocessing

Gambar 2, merupakan tahap preprossesing. Tujuannya adalah untuk memastikan bahwa data tersebut sesuai untuk analisis atau pelatihan model.

3.3 Implementasi LSTM

Pada tahap ini, dilakukan implementasi metode LSTM untuk klasifikasi teks berita breaking news. Ini mencakup pembuatan model LSTM, konfigurasi parameter, dan pelatihan model menggunakan dataset yang telah disiapkan.

3.4 Pengujian

Setelah model LSTM diimplementasikan, tahap pengujian dilakukan menggunakan dataset uji yang terpisah. Pengujian ini melibatkan inputkan data uji ke dalam model dan mengevaluasi kinerja Model ini menggunakan metrik seperti presisi, akurasi, perolehan, dan skor F1.

3.5 Hasil Evaluasi

Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan nilai sebenarnya materi tes. Hasil evaluasi ini memberikan pemahaman komprehensif tentang sejauh mana model LSTM dapat mengklasifikasikan berita secara akurat. Dengan menggunakan metrik-metrik ini, peneliti dapat menilai kehandalan model dan membuat penyesuaian jika diperlukan untuk meningkatkan kinerja model.

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Sampel}} \quad (1)$$

$$\text{Persisi} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{Recall} = \frac{TP}{FN + TP} \quad (3)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

Keterangan :

TP: Benar Positif.

TN: Benar Negatif.

FP: Positif Palsu.

FN : Negatif Palsu.

4. HASIL DAN PEMBAHASAN

Hasil model LSTM. Pada tahap ini spesifikasi dari sebuah model jaringan saraf tiruan (neural network) disusun secara berurutan (sequential).

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 200, 32)	320032
lstm (LSTM)	(None, 100)	53200
dense (Dense)	(None, 4)	404

Total params: 373636 (1.43 MB)
 Trainable params: 373636 (1.43 MB)
 Non-trainable params: 0 (0.00 Byte)

Gambar 3. Hasil Model LSTM

Gambar 3, merupakan lapisan LSTM (Long-Term Memory) yang sangat berguna untuk memproses rangkaian data seperti teks. LSTM memiliki keunggulan dalam memahami hubungan temporal data, dan model ini memiliki 100 unit LSTM. Setiap masukan string teks menghasilkan keluaran berupa vektor dengan panjang 100, dan parameter LSTM disesuaikan selama pelatihan untuk memahami pola temporal dalam data. Terakhir, ada lapisan padat yang bertindak sebagai lapisan keluaran model. Dalam contoh ini, terdapat 4 neuron pada lapisan padat, yang menunjukkan bahwa model harus diklasifikasikan menjadi 4 kelas berbeda. Jumlah parameter model adalah 373.636, yang merupakan jumlah dari semua parameter level (termasuk bobot dan perpindahan). Semua parameter model dapat disesuaikan selama proses pelatihan untuk meningkatkan kinerja model dalam tugas klasifikasi teks tertentu.

Tahap Pengujian :

Fase pengujian ini dilakukan setelah melatih model menggunakan data pelatihan dan sebelum menerapkan model di lingkungan produksi atau dunia nyata.

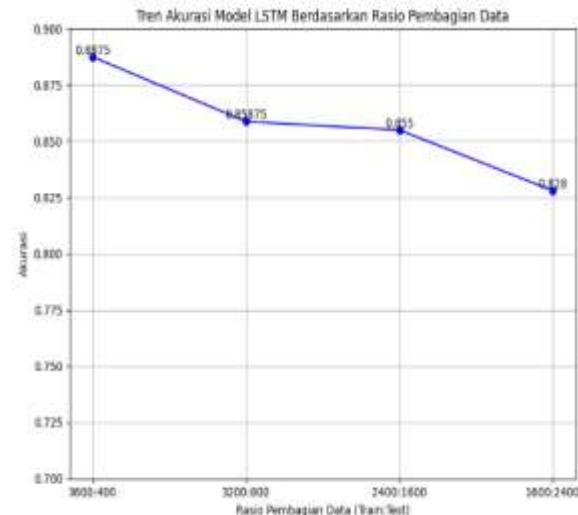
Tabel 1. Hasil Pengujian

Train:Test	Akurasi	Precision	Recall	F1-Score
3600:400	88.75%	88.79%	88.75%	88.76%
3200:800	85.875%	86.39%	85.88%	85.99%
2400:1600	85.5%	85.76%	85.5%	85.59%
1600:2400	82.8%	83.80%	82.83%	82.98%

Tabel 1, Dari hasil pengujian yang disajikan dalam tabel, terlihat bahwa ketika perbandingan jumlah data latih dan data uji semakin tidak seimbang, seperti pada kasus 3600:400 atau 3200:800, terjadi penurunan performa model, yang menunjukkan kemungkinan terjadinya overfitting.

Overfitting terjadi ketika model terlalu kompleks dan terlalu menyesuaikan dengan data latih, sehingga tidak mampu menggeneralisasi dengan baik pada data uji yang belum pernah dilihat sebelumnya. Untuk mengatasi overfitting pada model LSTM, beberapa strategi yang dapat dilakukan antara lain adalah penambahan data, pengurangan kompleksitas model, penyesuaian hyperparameter, dan penggunaan cross-validation untuk memvalidasi performa model secara lebih menyeluruh.

Hasil Perbandingan :



Gambar 4. Hasil Perbandingan

Gambar 4 Pada grafik diatas menunjukkan tren akurasi model LSTM berdasarkan rasio pembagian data latih dan uji. Grafik menampilkan rasio pembagian data pada sumbu x dan akurasi model pada sumbu y. Dari grafik, terlihat bahwa akurasi cenderung menurun saat ukuran data latih berkurang atau saat data uji menjadi lebih besar dibandingkan data latih. Ini menunjukkan bahwa semakin sedikit data yang digunakan untuk melatih model, semakin sulit bagi model untuk memprediksi data uji dengan benar. Namun, meskipun terjadi variasi dalam ukuran data latih, grafik menunjukkan bahwa model masih mempertahankan tingkat akurasi yang relatif stabil. Hasil akurasi yang diamati adalah 88.75% untuk pembagian data 3600:400, 85.875% untuk pembagian data 3200:800, 85.5% untuk pembagian data 2400:1600, dan 82.8% untuk pembagian data 1600:2400. Hal ini menunjukkan bahwa model LSTM tetap memberikan kinerja yang kuat dengan variasi pembagian data yang berbeda.

5. KESIMPULAN DAN SARAN

Klasifikasi teks berita terkini di Manggarai berdasarkan hasil penelitian yang diperoleh dengan metode LSTM, dapat disimpulkan bahwa proses penelitian ini menghasilkan klasifikasi teks berita yang efektif. Dengan dataset berjumlah 4000 data, di mana setiap kategori memiliki 1000 dataset,

dilakukan variasi pembagian data pada data train dan test dengan beberapa model, yaitu 3600:400, 3200:800, 2400:1600, dan 1600:2400. Model LSTM berhasil menunjukkan performa yang unggul dibandingkan dengan model lainnya, dengan akurasi sebesar 88.75%, precision 88.79%, recall 88.75%, dan F1-score 88.76% pada model pertama. Hasil ini menegaskan bahwa model mampu dengan baik dalam memahami dan mengelompokkan teks berita ke dalam kategori-kategori yang telah ditentukan. Dengan demikian, penelitian ini menunjukkan potensi yang kuat untuk penerapan metode LSTM dalam klasifikasi teks berita, dengan hasil yang menjanjikan dalam memprediksi kategori-kategori berita tertentu. Oleh karena itu untuk penelitian selanjutnya perlu dilakau evaluasi lebih lanjut terhadap teknik pengembangan model, penggunaan sumber data tambahan, penanganan ketidakseimbangan kelas.

DAFTAR PUSTAKA

- [1] T. Pustaka, "Tahun Publikasi," Vol. 7, No. 1, Pp. 681–686, 2023.
- [2] Yudi Widhiyasana, Transmissia Semiawan, Ilham Gibran Achmad Mudzakir, And Muhammad Randi Noor, "Penerapan Convolutional Long Short-Term Memory Untuk Klasifikasi Teks Berita Bahasa Indonesia," *J. Nas. Tek. Elektro Dan Teknol. Inf.*, Vol. 10, No. 4, Pp. 354–361, 2021, Doi: 10.22146/Jnteti.V10i4.2438.
- [3] N. Yuliyanti, "Implementasi Metode Long Short-Term Memory (Lstm) Pada Klasifikasi Ulasan Aplikasi Mobile Jkn," *Angew. Chemie Int. Ed.* 6(11), 951–952., Pp. 5–24, 2017, [Online]. Available: [Http://Repo.Iain-Tulungagung.Ac.Id/5510/5/Bab 2.Pdf](http://Repo.Iain-Tulungagung.Ac.Id/5510/5/Bab%202.Pdf)
- [4] H. Jelodar, Y. Wang, R. Orji, And S. Huang, "Deep Sentiment Classification And Topic Discovery On Novel Coronavirus Or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach," *IEEE J. Biomed. Heal. Informatics*, Vol. 24, No. 10, Pp. 2733–2742, 2020, Doi: 10.1109/JBHI.2020.3001216.
- [5] N. P. S. Wati And C. Pramatha, "Penerapan Long Short Term Memory Dalam Mengklasifikasi Jenis Ujaran Kebencian Pada Tweet Bahasa Indonesia," *J. Nas. Teknol. Inf. Dan Apl.*, Vol. 1, No. 1, Pp. 755–762, 2022.
- [6] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Adriansyah, M. H. Kafabi, And S. Sudianto, "Klasifikasi Judul Berita Clickbait Menggunakan RNN-LSTM," *J. Inform. J. Pengemb. IT*, Vol. 7, No. 2, Pp. 85–89, 2022, Doi: 10.30591/Jpit.V7i2.3401.
- [7] H. Rhomadhona And J. Permadi, "Klasifikasi Berita Kriminal Menggunakan Naïve Bayes Classifier (NBC) Dengan Pengujian K-Fold Cross Validation," *J. Sains Dan Inform.*, Vol.

- 5, No. 2, Pp. 108–117, 2019, Doi: 10.34128/Jsi.V5i2.177.
- [8] Pranowo, “Kearifan Lokal Jawa Menuju Konteks Global: Studi Makna Pragmatik,” *Salingka Maj. Ilm. Bhs. Dan Sastra*, Vol. 17, No. 2, Pp. 113–125, 2020.
- [9] Aji Priyambodo And Prihati Prihati, “Evaluasi Ekstraksi Fitur Klasifikasi Teks Untuk Peningkatan Akurasi Klasifikasi Menggunakan Naive Bayes,” *Elkom J. Elektron. Dan Komput.*, Vol. 13, No. 1, Pp. 159–175, 2020, Doi: 10.51903/Elkom.V13i1.277.
- [10] R. Nanda, E. Haerani, S. K. Gusti, And S. Ramadhani, “Klasifikasi Berita Menggunakan Metode Support Vector Machine,” *J. Nas. Komputasi Dan Teknol. Inf.*, Vol. 5, No. 2, Pp. 269–278, 2022, Doi: 10.32672/Jnkti.V5i2.4193.
- [11] S. Khairunnisa, A. Adiwijaya, And S. Al Faraby, “Pengaruh Text Preprocessing Terhadap Analisis Sentimen Komentar Masyarakat Pada Media Sosial Twitter (Studi Kasus Pandemi COVID-19),” *J. Media Inform. Budidarma*, Vol. 5, No. 2, P. 406, 2021, Doi: 10.30865/Mib.V5i2.2835.
- [12] B. A. H. Kholifatullah And A. Prihanto, “Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech,” *J. Informatics Comput. Sci.*, Vol. 04, Pp. 292–297, 2023, Doi: 10.26740/Jinacs.V4n03.P292-297.
- [13] T. Bastian Sianturi, I. Cholissodin, And N. Yudistira, “Penerapan Algoritma Long Short-Term Memory (LSTM) Berbasis Multi Fungsi Aktivasi Terbobot Dalam Prediksi Harga Ethereum,” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 7, No. 3, Pp. 1101–1107, 2023, [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [14] K. D. Larasati And A. H. Primandari, “Forecasting Bitcoin Price Based On Blockchain Information Using Long-Short Term Method,” *Param. J. Stat.*, Vol. 1, No. 1, Pp. 1–6, 2021, Doi: 10.22487/27765660.2021.V1.I1.15389.
- [15] Y. Tang And R. Salakhutdinov, “Learning Stochastic Feedforward Neural Networks,” *Adv. Neural Inf. Process. Syst.*, Pp. 1–9, 2013.