Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

Implementasi Bi-LSTM Untuk Klasifikasi Sembilan Kategori Teks Berita Bahasa Indonesia

Ali Imron, Erna Daniati[2,*], Dwi Harini[3]
Fakultas Teknik dan Ilmu Komputer, Program Studi Sistem Informasi, Universitas Nusantara PGRI Kediri, Kota Kediri, Indonesia
Email: [1]aliimron14031990@gmail.com, [2,*]ernadaniati@unpkediri.ac.id, [3]dwiharini@unpkediri.ac.id
Email Penulis Korespondensi: ernadaniati@unpkediri.ac.id

Abstrak−Penelitian ini bertujuan mengembangkan model klasifikasi teks berita berbahasa Indonesia ke dalam sembilan kategori menggunakan metode Bidirectional Long Short-Term Memory (Bi-LSTM). Berbeda dengan penelitian terdahulu yang mayoritas menggunakan dataset seimbang atau berfokus pada klasifikasi biner, riset ini menguji ketangguhan model pada kondisi ketidakseimbangan data (class imbalance) ekstrem di dunia nyata, dengan disparitas volume data antarkelas mencapai lebih dari tiga belas kali lipat. Tahapan penelitian meliputi preprocessing teks (case folding, punctuation removal, stemming Sastrawi), tokenisasi, padding, serta pembagian dataset menggunakan rasio 80:20. Arsitektur model mengintegrasikan embedding layer, Bi-LSTM, dropout, dan dense layer yang dioptimasi menggunakan teknik cost-sensitive class weight dan optimizer Adam. Hasil pengujian menunjukkan model memperoleh tingkat akurasi sebesar 0.85 dengan nilai weighted average precision, recall, dan f1-score masing-masing sebesar 0.85. Keunggulan utama metode ini terletak pada kemampuan ekstraksi fitur dua arah (forward dan backward) yang terbukti andal dalam memitigasi bias kelas mayoritas serta mereduksi ambiguitas semantik pada kategori minoritas. Kontribusi ilmiah dari riset ini memberikan pembuktian empiris dan referensi komprehensif mengenai efektivitas pemahaman konteks dwiarah dalam menjaga stabilitas performa klasifikasi teks multi-kelas pada lanskap data dunia nyata yang tidak ideal.

Kata Kunci: Bi-LSTM, Class imbalance, Klasifikasi teks berita, Natural Language Processing, Preprocessing teks

Abstract−This study aims to develop a classification model for Indonesian news texts across nine categories using the Bidirectional Long Short-Term Memory (Bi-LSTM) method. Unlike previous studies that predominantly utilized balanced datasets or focused on binary classification, this research evaluates the model's robustness under extreme real-world class imbalance, where the disparity in data volume between classes reaches over thirteen-fold. The research methodology encompasses text preprocessing (case folding, punctuation removal, and Sastrawi stemming), tokenization, padding, and dataset splitting at an 80:20 ratio. The model architecture integrates an embedding layer, a Bi-LSTM layer, a dropout layer, and a dense layer, which are optimized using a cost-sensitive class weight technique and the Adam optimizer. Experimental results demonstrate that the model achieved an accuracy of 0.85, with weighted average precision, recall, and F1-score values all at 0.85. The primary advantage of this method lies in its bidirectional feature extraction capability (forward and backward), which proves reliable in mitigating majority class bias and reducing semantic ambiguity within minority categories. The scientific contribution of this study provides empirical evidence and a comprehensive reference regarding the effectiveness of bidirectional contextual understanding in maintaining the performance stability of multi-class text classification within non-ideal, real-world data landscapes.

Keywords: Bi-LSTM, Class imbalance, News text classification, Natural Language Processing, Text preprocessing

1. PENDAHULUAN

Kemajuan teknologi digital telah menyebabkan peningkatan penyebaran berita berbahasa Indonesia melalui berbagai media, seperti portal berita online, media sosial, hingga aplikasi berbasis informasi digital. Banyaknya berita yang diproduksi dan dipublikasikan setiap hari menciptakan kebutuhan terhadap sistem otomatis yang dapat mengategorikan berita ke dalam kelompok tertentu sehingga proses pencarian, pengelolaan, dan pemahaman informasi menjadi lebih efisien [1]. Dalam hal ini, Natural Language Processing (NLP) atau pemrosesan bahasa alami memegang peranan penting karena teknologi tersebut memungkinkan komputer untuk menganalisis, memahami, serta menginterpretasikan bahasa manusia, termasuk konteks dan makna yang terdapat pada teks berita.

Meskipun demikian, penerapan NLP pada bahasa Indonesia masih menghadapi berbagai tantangan. Bahasa Indonesia memiliki karakteristik linguistik yang kompleks, seperti variasi morfologi, keragaman gaya penulisan, serta ambiguitas makna yang dapat memengaruhi performa model klasifikasi teks. Kondisi tersebut menyebabkan metode klasifikasi tradisional, seperti Naïve Bayes dan Support Vector Machine (SVM), belum mampu menghasilkan performa klasifikasi yang optimal pada berbagai kasus teks berbahasa Indonesia [2]. Sehingga diperlukan pendekatan yang lebih adaptif dalam memahami konteks kalimat secara menyeluruh.

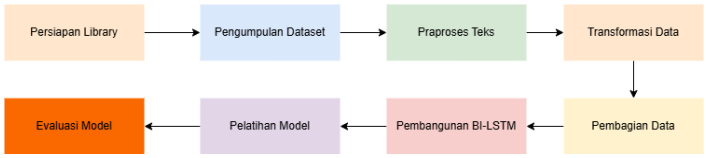
Salah satu metode yang saat ini banyak digunakan dalam klasifikasi teks adalah Bidirectional Long Short-Term Memory (Bi-LSTM). Model ini merupakan pengembangan dari Long Short-Term Memory (LSTM) yang bertujuan memperbaiki keterbatasan pada LSTM tradisional dengan melakukan pemrosesan urutan kata dari dua arah sekaligus, yaitu dari awal ke akhir (forward) dan dari akhir ke awal (backward) [3], [4]. Pendekatan dua arah tersebut membuat model mampu memahami keterkaitan antar kata secara lebih mendalam dan kontekstual. Oleh karena itu, Bi-LSTM dianggap lebih unggul dalam menangkap makna keseluruhan suatu teks dibandingkan metode konvensional yang hanya memproses data secara satu arah dan berurutan.

Berdasarkan beberapa studi terdahulu, implementasi metode deep learning seperti LSTM, Bi-LSTM, dan GRU terbukti efektif untuk klasifikasi teks berbahasa Indonesia [5], [6]. Meskipun demikian, mayoritas penelitian tersebut masih mengasumsikan distribusi data yang ideal baik melalui penggunaan dataset yang seimbang (balanced dataset) maupun pembatasan pada klasifikasi biner dengan jumlah kategori label yang sedikit. Pada realitasnya, ketika cakupan klasifikasi

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

diperluas menjadi skala multi-kelas yang lebih kompleks, karakteristik data berita digital di Indonesia sangat dinamis dan memicu tantangan teknis baru yang belum banyak dieksplorasi secara mendalam [7], [8].
Tantangan krusial yang muncul pada klasifikasi multi-kelas berskala besar adalah masalah ketidakseimbangan data (class imbalance) yang ekstrem. Skenario klasifikasi sembilan kategori berita berbahasa Indonesia (news, hot, finance, travel, inet, health, oto, food, sport) dalam lanskap media digital melahirkan disparitas kuantitas sampel yang sangat mencolok antarkelas. Kategori umum seperti news cenderung mendominasi secara masif sebagai kelas mayoritas, sementara kategori spesifik seperti sport bertindak sebagai kelas minoritas dengan volume data yang sangat terbatas hingga membentuk ketimpangan lebih dari tiga belas kali lipat. Jika kondisi ini diabaikan, bias algoritma akan terjadi selama proses pembelajaran; model deep learning konvensional akan mengalami overfitting pada pola kelompok dominan dan mengabaikan karakteristik data minoritas, sehingga secara drastis menurunkan akurasi prediksi untuk kategori-kategori kecil [9], [10]. Celah permasalahan teknis mengenai degradasi performa akibat class imbalance inilah yang menjadi hambatan utama dalam mewujudkan sistem klasifikasi otomatis yang adil (fair) dan akurat.
Untuk menyelesaikan permasalahan spesifik tersebut, penelitian ini mengajukan penerapan metode Bidirectional Long Short-Term Memory (Bi-LSTM). Kebaruan dari penelitian ini terletak pada evaluasi empiris mengenai ketangguhan ekstraksi fitur kontekstual dua arah dalam mereduksi ambiguitas semantik pada kondisi class imbalance ekstrem berskala sembilan kategori. Lapisan Bi-LSTM diandalkan secara penuh untuk menangkap dependensi kata dari arah maju (forward) maupun mundur (backward) guna membedakan tumpang tindih makna kosakata antarkelas yang timpang [11]. Adapun teknik class weight diaplikasikan dalam penelitian ini murni sebagai langkah preventif standar (preventive best practice) pada fase optimasi untuk menjaga stabilitas pelatihan jaringan saraf dari risiko bias kelas mayoritas sejak awal, sejalan dengan pendekatan optimasi fungsi kerugian (loss optimization) pada studi pemrosesan bahasa alami modern [12], [13].
Dengan demikian, tujuan utama penelitian ini adalah mengembangkan dan mengevaluasi keandalan arsitektur Bi-LSTM dalam mengklasifikasikan sembilan kategori berita berbahasa Indonesia yang memiliki karakteristik sebaran data tidak seimbang. Model yang dikembangkan akan dievaluasi menggunakan classification report dan confusion matrix untuk mengukur kemampuan model dalam membedakan setiap kategori berita secara akurat [14]. Kontribusi ilmiah dari riset ini terletak pada penyajian analisis mengenai bagaimana pemahaman konteks dua arah mampu mempertahankan performa klasifikasi teks di tengah tantangan disparitas volume data yang masif, sehingga dapat menjadi referensi komprehensif bagi pengembangan sistem NLP pada skenario dunia nyata yang tidak ideal.

2. METODOLOGI PENELITIAN



Gambar 1. Alur Penelitian

Alur kerja dalam studi ini disusun melalui tahapan terstruktur guna menjamin setiap tingkatan. Mulai dari tata kelola data hingga analisis akhir berjalan secara efektif dan menghasilkan luaran yang valid. Investigasi ini menerapkan metode eksperimental berbasis deep learning yang dieksekusi menggunakan ekosistem Python serta framework TensorFlow. Secara sistematis, rangkaian eksperimen dipecah menjadi beberapa tahapan integral, yang diawali dengan konfigurasi lingkungan perangkat lunak serta pustaka pendukung. Selanjutnya, proses dilanjutkan pada tahap manipulasi awal data (preprocessing), perancangan topologi model, fase training, hingga diakhiri dengan verifikasi serta penilaian performa guna mengukur akurasi model dalam menyelesaikan komputasi yang ditargetkan.

2.1 Persiapan Library

Tahap awal penelitian dilakukan dengan menyiapkan lingkungan kerja berbasis Google Colab serta menginstal library yang dibutuhkan seperti TensorFlow, NLTK, NumPy, Pandas, dan Scikit-learn. Selain itu, dilakukan proses mounting Google Drive sebagai media penyimpanan dataset. Pada tahap ini juga dilakukan pengunduhan resource tambahan dari NLTK seperti stopwords, punkt, dan wordnet yang digunakan dalam proses preprocessing teks [15]. Tahapan ini penting untuk memastikan seluruh dependensi sistem dapat berjalan dengan baik sebelum memasuki proses pengolahan data.

2.2 Pengumpulan Dataset

Riset memanfaatkan Indonesian News Title Dataset dari platform Kaggle berformat CSV. Untuk efisiensi komputasi, proses ekstraksi fitur secara ketat hanya mengisolasi atribut title sebagai variabel prediktor dan category sebagai label target. Entitas metadata seperti date dan url direduksi karena tidak relevan dengan tujuan klasifikasi. Tahap prapemrosesan awal juga mencakup purifikasi data dari anomali (nilai kosong dan redundansi) guna menjamin validitas representasi informasi sebelum diumpankan ke dalam model. Diagnostik terhadap sembilan kategori berita mengungkap adanya anomali class imbalance yang sangat ekstrem.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

Tabel 1. Identifikasi Data Tidak Seimbang (Class Imbalance)

Aspek Analisis	Hasil
Total Dataset	87.017 data
Kelas Mayoritas	News (32.360 data)
Kelas Minoritas	Sport (2.436 data)
Indikasi Class Imbalance	Ya
Solusi yang Diterapkan	Class Weight

Merujuk pada Tabel 1, kelas mayoritas (news) mendominasi secara masif dengan 32.360 entitas, berbanding jauh dengan kelas minoritas (sport) yang hanya memiliki 2.436 entitas. Ketimpangan ini sangat rentan memicu bias algoritma, di mana model berisiko mengalami overfitting pada kelas yang mendominasi dan gagal mengenali pola kelas minoritas. Sebagai intervensi metodologis, arsitektur Bi-LSTM dikalibrasi menggunakan teknik class weight. Mekanisme ini bekerja dengan mendistribusikan penalti optimasi yang jauh lebih berat terhadap kesalahan klasifikasi pada kelompok data minoritas. Kompensasi algoritmik ini secara efektif memaksa model untuk mengekstraksi parameter dari seluruh kelas secara ekuivalen, sehingga meminimalisasi bias mayoritas sekaligus mendongkrak kemampuan generalisasi prediksi.

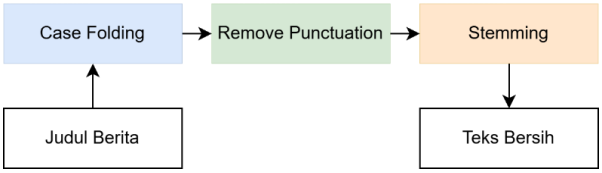
	title	category
0	kemnaker awas tka di meikarta	finance
1	bni digital bni java jazz 2020	finance
2	terbang ke australia edhy prabowo mau genjot b...	finance
3	ojk siap stimulus ekonomi antisipasi dampak co...	finance
4	saran buat anies rk yang mangkir rapat banjir ...	finance
...	...	...
91012	tumpah air panas di pesawat kamu bisa tuntut m...	travel
91013	foto bal 9 destinasi paling instagramable tahu...	travel
91014	game bikin turis ini libur ke jepang untuk car...	travel
91015	keluarga depak dari pesawat maskapai bilang ba...	travel
91016	kapal raib di segitiga muda nyaris abad baru k...	travel

91017 rows × 2 columns

Gambar 2. Dataset Berita 9 Kategori

2.3 Text Preprocessing

Prapemrosesan merupakan fase krusial dalam arsitektur Natural Language Processing (NLP) yang secara fundamental menentukan kualitas representasi data masukan bagi model klasifikasi. Tahap preprocessing berperan langsung dalam menentukan kualitas data masukan (input) yang akan diproses oleh model [16]. Implementasi prapemrosesan yang komprehensif terbukti mampu meningkatkan akurasi prediktif secara signifikan [17], [18]. Dalam konteks data berita digital, informasi mentah sering kali mengandung noise berupa ketidakkonsistenan huruf kapital, keberadaan simbol non-alfanumerik, serta variasi morfologis kata yang berpotensi memicu redundansi semantik [19]. Kegagalan dalam menangani variasi ini menyebabkan sparse feature space, yang secara langsung menurunkan efisiensi pembelajaran model [20]. Oleh karena itu, penelitian ini mengintegrasikan serangkaian prosedur prapemrosesan sistematis meliputi case folding, punctuation removal, dan stemming untuk menstandarisasi kolom title sebagai fitur utama. Alur operasional prapemrosesan teks pada riset ini digambarkan secara visual dalam Gambar 3.



Gambar 3. Alur Pre-Processing Teks

Prosedur diawali dengan case folding untuk melakukan normalisasi karakter menjadi huruf kecil, yang krusial untuk memastikan bahwa variasi kapitalisasi tidak diperlakukan sebagai entitas linguistik yang berbeda. Selanjutnya, metode regular expression diaplikasikan untuk eliminasi tanda baca, guna membuang simbol yang tidak berkontribusi pada fitur semantik. Langkah terakhir adalah stemming berbasis pustaka Sastrawi untuk mereduksi kata berimbuhan menjadi bentuk dasar (misalnya, konversi "berjualan" dan "penjualan" menjadi "jual"). Secara teoretis, proses normalisasi morfologis ini sangat esensial dalam pengolahan bahasa Indonesia untuk meningkatkan kepadatan semantik dalam vektor fitur [21], [22]. Dengan menyederhanakan representasi leksikal melalui prosedur ini, model Bi-LSTM dapat memfokuskan ekstraksi

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

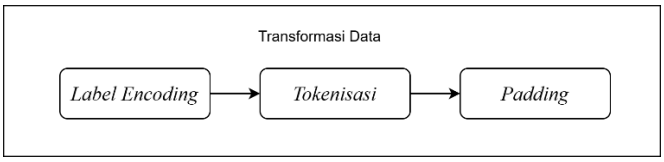
fiturnya pada dependensi kontekstual yang lebih relevan, sehingga meningkatkan objektivitas dan konsistensi klasifikasi pada data yang belum pernah dilihat sebelumnya (unseen data).

```
def preprocessing_text(text):
    text = text.strip()
    text = text.lower()
    text = rem_punc(text)
    text = stem_text(text)
    return text
```

Gambar 4. Source Code Pre-Processing Data

2.4 Transformasi Data

Transformasi Representasi DataTahapan transformasi merupakan fondasi komputasional untuk mengonversi data tekstual ke dalam bentuk vektor numerik yang dapat diproses oleh arsitektur deep learning [23]. Mengingat jaringan saraf tiruan beroperasi melalui kalkulasi matriks, data mentah harus melalui serangkaian prosedur normalisasi yang ketat agar karakteristik semantik teks dapat diinterpretasikan secara matematis.



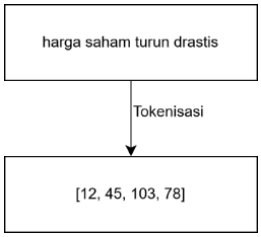
Gambar 5. Alur Transformasi Data

2.4.1 Label Encoding

Fase ini mengonversi label kategori teks menjadi representasi integer unik. Sebanyak sembilan kategori berita dipetakan ke dalam skalar numerik [0, 8]. Penggunaan label encoding bertujuan untuk memfasilitasi optimasi fungsi sparse categorical crossentropy sebagai mekanisme pengukur loss selama propagasi balik. Penting untuk dicatat bahwa pemetaan ini bersifat nominal tidak memiliki relasi hierarkisehingga setiap kelas diperlakukan sebagai entitas diskret yang independen dalam klasifikasi multikelas.

2.4.2 Tokenisasi dan Vektorisasi

Tokenisasi berfungsi membedah struktur leksikal kalimat menjadi unit-unit token dasar. Peneliti menggunakan kelas Tokenizer dari TensorFlow Keras untuk mengonstruksi kamus kosakata (vocabulary) berdasarkan frekuensi kemunculan term dalam korpus. Guna menekan kompleksitas komputasi dan meminimalisasi overfitting, batasan kosakata ditetapkan pada parameter vocab_size=20.000. Untuk menangani variasi leksikal di luar korpus pelatihan, diterapkan token khusus Out of Vocabulary (OOV). Integrasi antara indeks numerik unik dengan urutan kata memungkinkan model Bi-LSTM mengekstraksi dependensi kontekstual secara lebih presisi daripada sekadar pemrosesan karakter mentah [24], [25].

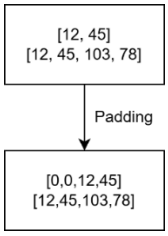


Gambar 6. Transformasi Teks Menjadi Representasi Numerik (Tokenisasi)

2.4.3 Padding Sequence

Karena setiap judul berita memiliki panjang sintaksis yang beragam, diperlukan prosedur padding untuk menstandarisasi dimensi input. Riset ini menetapkan max_length=100 dengan mekanisme post-padding menggunakan nilai nol (0) sebagai filler bagi sekuens yang kurang panjang, serta teknik truncating='post' bagi sekuens yang melebihi batas. Standarisasi dimensi ini mutlak diperlukan untuk memungkinkan pemrosesan data dalam format matriks terstruktur, yang secara signifikan meningkatkan efisiensi komputasi pada GPU selama proses batch training [26].

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS



Gambar 7. Penyamakan Panjangnya Sequence (Teknik Padding)

2.5 Hyperparameter & Pembagian Dataset

Menjelang fase komputasi arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM), inisialisasi hyperparameter merupakan tahapan fundamental yang mendikte trajektori pembelajaran jaringan saraf tiruan. Berbeda dengan bobot model yang diperbarui secara otomatis, hyperparameter merupakan variabel kontrol statis yang harus dikalibrasi oleh peneliti secara empiris guna mencapai keseimbangan antara akurasi prediktif dan efisiensi komputasi [27]. Kompleksitas penyesuaian parameter ini secara langsung menentukan stabilitas proses training dan kecepatan konvergensi algoritma. Rincian komprehensif mengenai matriks konfigurasi eksperimental yang ditetapkan dalam penelitian ini didokumentasikan secara terperinci pada Tabel 2.

Tabel 2. Hyperparameter Model

Hyperparameter	Nilai
Vocabulary Size	20.000
Embedding Dimension	128
Max Length	100
OOV Token	<OOV>
Truncating Type	post
Jumlah Unit Bi-LSTM	64
Dropout Rate	0.5
Dense Layer	64 neuron
Output Layer	9 neuron
Fungsi Aktivasi Dense	ReLU
Fungsi Aktivasi Output	Softmax
Optimizer	Adam
Loss Function	Sparse Categorical Crossentropy
Epoch	20
Batch Size	128
Train-Test Split	80% : 20%

Konfigurasi ini dirancang khusus untuk meminimalkan beban memori pada parameter vocabulary size dan embedding dimension, sekaligus menstandarisasi matriks input melalui penetapan max_length yang ekuivalen di seluruh sekuens berita. Pada aspek arsitektural, lapisan Bi-LSTM diintegrasikan dengan 64 unit sel memori guna memaksimalkan ekstraksi probabilitas kondisional kontekstual dari dua arah secara simultan. Guna mereduksi potensi overfitting yang umum terjadi pada pemrosesan dimensi data yang masif, mekanisme regulasi diterapkan melalui parameter dropout sebesar 0.5, yang berfungsi menonaktifkan probabilitas aktivasi neuron secara acak [28]. Tahap optimasi pembelajaran dieksekusi menggunakan Adaptive Moment Estimation (Adam) untuk pembaruan learning rate secara dinamis, dikombinasikan dengan fungsi kerugian sparse categorical crossentropy yang secara spesifik dirancang untuk permasalahan pemetaan kelas multikategori.

Penentuan kuantitas epoch dan batch size turut diposisikan secara proporsional untuk mencegah model terjebak dalam dilema bias-varians (bias-variance tradeoff), memastikan bahwa sistem mampu mengekstraksi dependensi semantik secara holistik tanpa menghafal anomali (noise) dari data laten. Untuk memvalidasi kapabilitas generalisasi model secara objektif, implementasi korpus dilanjutkan dengan prosedur partisi data (data splitting) menjadi dua domain independen. Pemisahan dieksekusi dengan mendistribusikan 80% himpunan data sebagai instrumen pelatihan adaptasi pola fitur (training set), sedangkan 20% sisanya diisolasi sebagai instrumen pengujian performa prediksi (testing set). Rasio 80:20

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

ini merupakan protokol empiris yang secara luas diakui dalam literatur machine learning modern untuk menjaga integritas evaluasi algoritma [27]. Operasi pemisahan ini diimplementasikan memanfaatkan fungsi komputasional train_test_split dari pustaka scikit-learn, sebagaimana diilustrasikan secara prosedural pada Gambar 8 Penentuan parameter random_state diinjeksikan dalam skrip tersebut untuk mengeliminasi bias acak, sehingga arsitektur pembagian data bersifat deterministik dan menjamin bahwa serangkaian hasil komputasi dari model ini dapat direproduksi secara konsisten (reproducibility) pada eksperimen di masa mendatang.

```
# Bagi dataset menjadi data latih 0.8 dan data tes 0.2
x_train, x_test, y_train, y_test = train_test_split(
    x, y,
    test_size=0.2,
    random_state=42
)
```

Gambar 8. Source Code Pembagian Dataset

2.6 Model Bidirectional-LSTM

Desain arsitektur komputasional dalam penelitian ini diinisiasi oleh lapisan Embedding, yang mengeksekusi transformasi dimensi terhadap representasi token numerik diskret menjadi vektor kontinu berdensitas tinggi (dense vector). Transformasi spasial ini merupakan fondasi vital dalam pemrosesan bahasa alami karena indeks nominal biasa gagal merepresentasikan korelasi semantik. Melalui ruang proksimitas embedding, entitas leksikal yang berbagi kedekatan makna linguistik akan diproyeksikan ke titik koordinat spasial yang berdekatan, memungkinkan model untuk mengekstraksi dan mempelajari kedalaman relasi kata secara matematis. Representasi laten tersebut kemudian ditransmisikan menuju komponen inti arsitektur, yakni modul Bidirectional Long Short-Term Memory (Bi-LSTM). Modul ini mengoperasikan dua lintasan sel memori independen bergerak secara progresif (forward) dan regresif (backward) melintasi sekuens waktu yang keluaran fiturnya kemudian dikonjugasikan (concatenate). Mekanisme pemindaian dwiarah ini menjamin kapabilitas jaringan dalam mengasimilasi informasi kontekstual yang mendahului maupun mengikuti sebuah term, yang mana merupakan syarat mutlak untuk menekan ambiguitas semantik pada klasifikasi teks berita yang dinamis [29].

Tabel 3. Hyperparameter Model

No	Layer	Output Shape	Parameter	Fungsi
1	Embedding	(None, 100, 128)	2.560.000	Mengubah token menjadi vektor embedding berdimensi 128
2	Bidirectional LSTM (64 Unit)	(None, 128)	98.816	Menangkap konteks teks dari arah maju dan mundur
3	Dropout (0.5)	(None, 128)	0	Mengurangi overfitting dengan menonaktifkan neuron secara acak
4	Dense (ReLU)	(None, 64)	8.256	Mengolah fitur hasil ekstraksi dari Bi-LSTM
5	Dense (Softmax)	(None, 9)	585	Menghasilkan probabilitas untuk 9 kategori berita

Guna mereduksi probabilitas ko-adaptasi neuron yang berujung pada fenomena overfitting, topologi jaringan ini diintegrasikan dengan mekanisme regulasi Dropout dengan rasio penalti diatur pada 0.5. Intervensi stokastik ini mendisaktivasi separuh unit neuron secara acak pada setiap iterasi training, secara empiris memaksa arsitektur untuk mendistribusikan bobot pembelajaran secara ekuivalen dan mendongkrak ketangguhan generalisasi pada set pengujian. Sinyal fitur tingkat tinggi hasil ekstraksi Bi-LSTM kemudian diteruskan ke dalam infrastruktur fully connected layer (lapisan Dense) yang menampung 64 unit neuron. Lapisan tersembunyi ini difasilitasi oleh fungsi aktivasi Rectified Linear Unit (ReLU) guna memproses agregasi fitur non-linear sebelum klasifikasi final. Pengadopsian ReLU bukan tanpa alasan komputasional; aktivasi ini secara efektif meredam ancaman degradasi gradien (vanishing gradient problem) sekaligus mengakselerasi trajektori konvergensi iteratif dari model[30].
Sebagai konklusi arsitektural, lapisan Dense keluaran dikonfigurasi dengan 9 unit neuron untuk memetakan ruang prediksi ke dalam sembilan target klasifikasi berita. Lapisan terminal ini ditenagai oleh aktivasi Softmax, yang bertugas mengonversi akumulasi nilai logit menjadi distribusi metrik probabilitas. Klasifikasi absolut ditentukan berdasarkan nilai probabilitas multivariabel tertinggi. Secara holistik, sinkronisasi antara Embedding, Bi-LSTM, Dropout, Dense, dan Softmax ini membentuk kerangka kerja yang sangat tangguh dalam mengekskavasi pola linguistik laten. Visualisasi struktural dari sekuens lapisan komputasional ini dapat ditinjau pada Gambar 9.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

```
# Membangun model LSTM (BiDirectional) untuk klasifikasi teks (PERANCANGAN MODEL)
model = tf.keras.Sequential([
    tf.keras.layers.Embedding(vocab_size, 128, input_length=max_length),
    tf.keras.layers.Bidirectional(tf.keras.layers.LSTM(64)),
    tf.keras.layers.Dropout(0.5),
    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.Dense(9, activation='softmax')
])
```

Gambar 9. Source Code Model Bi-LSTM

2.7 Pelatihan Model

Fase krusial dalam komputasi prediktif ini berpusat pada eksekusi pelatihan arsitektur Bi-LSTM guna mengekstrak probabilitas korelasional antara struktur leksikal teks dan target kategori. Matriks input yang telah terstandarisasi melalui vektorisasi (tokenization) dan penyelarasan dimensi (padding) diinjeksikan sebagai variabel prediktor, sementara label encoding beroperasi sebagai konvensi target prediksi. Secara iteratif, jaringan melakukan kalibrasi parameter internal meliputi penyesuaian bobot sinaptik dan bias melalui mekanisme propagasi balik guna menekan tingkat galat (error rate) secara progresif. Secara operasional, iterasi komputasi ini diimplementasikan dengan memanggil rutin fit() pada pustaka TensorFlow Keras, yang dikonfigurasi menggunakan siklus 20 epoch dan batch size sebesar 128. Pemilihan ukuran batch 128 merupakan pendekatan empiris yang krusial untuk menjaga stabilitas trajektori gradien selama konvergensi stokastik, sekaligus meminimalkan latensi komputasi pada pemrosesan skala besar [31]. Secara simultan, dataset pengujian dialihfungsikan sebagai set validasi silang untuk mengevaluasi metrik performa secara objektif pada setiap akhir iterasi.

Tabel 4. Hyperparameter Model

Parameter	Nilai	Keterangan
Optimizer	Adam	Digunakan untuk memperbarui bobot model selama pelatihan
Loss Function	Sparse Categorical Crossentropy	Digunakan untuk klasifikasi multikelas dengan label numerik
Metrics	Accuracy	Digunakan untuk mengukur tingkat ketepatan prediksi model
Epoch	20	Jumlah maksimum siklus pelatihan
Batch Size	128	Jumlah data yang diproses dalam satu iterasi
Validation Data	Data Testing	Digunakan untuk memantau performa model selama training
Class Weight	Balanced	Menyeimbangkan kontribusi setiap kelas saat pelatihan
Callback	Custom Callback	Menghentikan training jika akurasi mencapai 95%
Target Accuracy Callback	0.95	Batas akurasi untuk menghentikan pelatihan secara otomatis

Untuk mengorkestrasi pembaruan bobot secara dinamis, penelitian ini mengadopsi algoritma Adaptive Moment Estimation (Adam). Konfigurasi optimizer ini terbukti secara saintifik sangat superior dalam arsitektur pembelajaran mendalam (deep learning) karena kapabilitasnya memodifikasi learning rate secara adaptif berdasarkan kalkulasi momen gradien, yang secara efektif mencegah model terperangkap pada kondisi minimum lokal (local minima)[32]. Selanjutnya, guna memitigasi anomali ketimpangan distribusi di mana entitas kelas dominan seperti news berpotensi mendistorsi prediksi terhadap kelas minoritas seperti sport penelitian ini mengimplementasikan protokol cost-sensitive learning berupa class weight. Memanfaatkan modul compute_class_weight() dari Scikit-Learn, algoritma memformulasikan penalti optimasi secara otomatis dan proporsional. Intervensi matematis ini memaksa model untuk mendistribusikan atensi pembelajaran secara ekuivalen pada seluruh spektrum target, tanpa memandang rasio kuantitas sampel aslinya [33]. Sebagai lini pertahanan terakhir terhadap risiko degradasi generalisasi algoritmik (overfitting), mekanisme regulasi callback diintegrasikan secara langsung ke dalam prosedur training, sebagaimana diilustrasikan secara struktural pada Gambar 10.

```
class myCallback(tf.keras.callbacks.Callback):
    def on_epoch_end(self, epoch, logs={}):
        if logs.get('accuracy') >= 0.95:
            print("\nstop")
            self.model.stop_training = True

callbacks = myCallback()
```

Gambar 10. Source Code Pelatihan Model

Protokol interupsi ini dikalibrasi untuk menghentikan fase pelatihan secara prematur (early stopping) segera setelah akurasi akusisi menyentuh ambang absolut 95%. Terminasi otomatis ini sangat esensial untuk mencegah model menghafal anomali sisa (overtraining) sekaligus mereduksi beban waktu komputasi secara signifikan. Konklusi dari sinkronisasi antara optimasi Adam, injeksi pembobotan kelas adaptif, dan regulasi callback ini menghasilkan arsitektur klasifikasi teks yang sangat stabil, efisien, dan tangguh (robust) terhadap variansi data yang tidak terduga. Sintaksis implementasi final dari eksekusi pelatihan ini menggunakan rutin fit dari TensorFlow dapat ditinjau secara komprehensif pada Gambar 11.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

```
# melatih model
history = model.fit(
    padded, y_train,
    epochs=20,
    batch_size=128,
    validation_data=(testing_padded, y_test),
    callbacks=[callbacks],
    class_weight=class_weights
)
```

Gambar 11. Source Code Pelatihan Model

2.8 Evaluasi Model

Tahap evaluasi model dilakukan untuk menilai kinerja model Bi-LSTM dalam mengklasifikasikan data teks secara tepat dan stabil. Proses evaluasi ini bertujuan mengukur kemampuan generalisasi model terhadap data yang tidak terlibat pada tahap pelatihan. Salah satu metrik yang digunakan adalah accuracy, yaitu perbandingan antara jumlah prediksi yang benar dengan keseluruhan data uji. Akan tetapi, akurasi saja belum mampu menggambarkan performa model secara menyeluruh, terutama ketika distribusi kelas pada dataset tidak seimbang. Oleh sebab itu, penelitian ini turut memanfaatkan classification report yang meliputi metrik precision, recall, dan f1-score untuk setiap kelas. Seluruh proses evaluasi diimplementasikan menggunakan pustaka scikit-learn.
Selain itu, penelitian ini juga memanfaatkan confusion matrix untuk memvisualisasikan perbandingan antara hasil prediksi model dan label sebenarnya. Melalui confusion matrix, analisis terhadap kesalahan klasifikasi dapat dilakukan secara lebih rinci, termasuk identifikasi false positive dan false negative pada setiap kategori. Visualisasi tersebut membantu dalam memahami pola kesalahan yang dihasilkan model selama proses klasifikasi berlangsung. Dengan penggunaan berbagai metrik evaluasi tersebut, performa model dapat dikaji secara lebih menyeluruh sehingga dapat menjadi dasar dalam menilai kelayakan model maupun menentukan pengembangan lanjutan yang diperlukan.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

3.1.1 Analisis Distribusi Dataset

Hasil Evaluasi terhadap proporsi data pra-pelatihan esensial dilakukan untuk memetakan sebaran sembilan kategori berita (news, hot, finance, travel, inet, health, oto, food, dan sport). Langkah krusial ini diambil demi mendeteksi asimetri informasi atau ketidakseimbangan kelas (class imbalance) sejak dini, karena ketimpangan jumlah sampel dapat mendistorsi kemampuan model dalam mengenali karakteristik unik dari setiap kelompok berita. Berdasarkan hasil pemetaan yang tersaji pada Tabel 5, terlihat disparitas kuantitas yang sangat mencolok antar-kategori. Sektor news mendominasi sebagai kelas mayoritas dengan volume mencapai 32.360 sampel, sedangkan sport memosisikan diri sebagai kelas minoritas dengan keterbatasan data sebanyak 2.436 sampel. Perbedaan yang menyentuh angka lebih dari tiga belas kali lipat ini mengonfirmasi adanya kendala class imbalance. Jika dibiarkan tanpa penanganan, bias algoritma akan terjadi; model cenderung mengalami over-fitting pada pola kelompok dominan dan mengabaikan karakteristik data minoritas, sehingga menurunkan akurasi prediksi untuk kategori-kategori kecil.

Tabel 5. Data Tidak Seimbang (Class Imbalance)

Aspek Analisis	Hasil
Total Dataset	87.017 data
Kelas Mayoritas	News (32.360 data)
Kelas Minoritas	Sport (2.436 data)
Indikasi Class Imbalance	Ya
Solusi yang Diterapkan	Class Weight

Guna mereduksi dampak negatif dari tidak meratanya sebaran data tersebut, riset ini mengintegrasikan fungsi class weight ke dalam siklus pelatihan Bi-LSTM. Mekanisme ini bekerja secara adaptif dengan mengalokasikan nilai penalti atau bobot kesalahan yang lebih tinggi pada kelompok data minim (sport), sembari memperkecil bobot pada sektor dengan pasokan data melimpah (news). Strategi pembobotan ini memaksa model untuk mempelajari seluruh variasi kategori secara berimbang, mengikis bias mayoritas, serta mendongkrak kapabilitas generalisasi sistem dalam mengklasifikasikan dokumen berita secara lebih objektif dan akurat.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

3.1.2 Hasil Pre-Processing

Sebelum memasuki fase komputasi arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM), data tekstual mentah wajib melewati skema normalisasi guna mentransformasikannya menjadi format yang steril dan terstandar. Pemangkasan variasi kebahasaan yang tidak berkontribusi pada makna ini krusial untuk mengoptimalkan penyerapan informasi esensial oleh model. Dalam eksperimen ini, rekayasa data tersebut mengintegrasikan tiga operasi fundamental: penyeragaman huruf (case folding), eliminasi karakter non-alfabet (remove punctuation), serta pemotongan afiksasi (stemming) berbasis modul Sastrawi. Rekam jejak perubahan transaksional data ini didokumentasikan secara rinci pada Tabel 6.

Tabel 6. Hasil Pre-Processing Teks

No	Sebelum Preprocessing	Sesudah Preprocessing
1	Posting Foto Gading Marten, DJ Wilda Dianggap ...	posting foto gading marten dj wilda anggap mak...
2	Pemerintah Hadang Ide Karantina di Kapal Perang	perintah hadang ide karantina di kapal perang
3	Servis Orang Baru Pulang dari Singapura, Tukan...	servis orang baru pulang dari singapura tukang...
4	Harga Emas Cetak Rekor Tertinggi Sepanjang Masa!	harga emas cetak rekor tinggi panjang masa
5	Tak Habis Diminum, Sisa Seduhan Kopi Bisa Dima...	tak habis minum sisa seduh kopi bisa manfaat u...

Merujuk pada paparan di Tabel 6, fungsionalitas case folding berhasil menyamakan persepsi linguistik sistem dengan mengonversi seluruh huruf kapital maupun kombinasi karakter menjadi bentuk lowercase. Sebagai ilustrasi fungsional, frasa "Posting Foto Gading Marten, DJ Wilda Dianggap..." ditransmutasikan menjadi "posting foto gading marten dj wilda anggap...". Langkah pengondisian ini memitigasi risiko pembengkakan dimensi kosakata (vocabulary size). Tanpa adanya penyusutan ortografis ini, entitas seperti "Pemerintah" dan "pemerintah" akan terbaca sebagai dua token independen yang berbeda oleh algoritma, sehingga menurunkan efisiensi pembelajaran model. Fase berikutnya berfokus pada aspek sterilisasi teks melalui pembersihan simbol (remove punctuation), di mana seluruh karakter di luar struktur alfanumerik disisihkan dari korpus. Implementasinya terlihat pada pelenyapan tanda koma pada untaian kalimat "Servis Orang Baru Pulang dari Singapura, Tukang..." serta tanda seru dari judul berita "Harga Emas Cetak Rekor Tertinggi Sepanjang Masa!". Eksklusi simbol-simbol ini bertujuan mengeliminasi derau (noise) yang hampa nilai informasi, sehingga lapisan pembelajar Bi-LSTM dapat berkonsentrasi penuh pada kata-kata bermuatan semantik esensial sekaligus menyederhanakan matriks representasi data.

Sebagai pemungkas skema praproses, penanganan reduksi morfologis atau stemming diaplikasikan untuk memetakan kata berimbuhan kembali pada leksem dasarnya. Berdasarkan data pengujian, efektivitas pustaka Sastrawi terbukti mampu menyusutkan kata "pemerintah" menjadi "perintah", "terbalik" menjadi "balik", "pengusaha" menjadi "usaha", "diminum" menjadi "minum", dan "seduhan" menjadi "seduh". Kompresi variasi infleksi ini mereduksi densitas istilah unik dalam korpus, mempermudah pemodelan dalam mengidentifikasi korelasi makna lintas dokumen, dan secara linier menyajikan input yang lebih matang untuk proses tokenisasi, padding, serta klasifikasi akhir.

3.1.3 Hasil Label Encoding

Arsitektur deep learning beroperasi murni melalui kalkulasi matematis. Oleh karena itu, label kategori berita yang awalnya berupa teks (string) wajib dikonversi menjadi angka melalui teknik label encoding agar algoritma dapat memprosesnya. Sebagaimana didokumentasikan pada Tabel 7, sembilan kelas berita (dari news hingga sport) dipetakan secara konsisten menjadi bilangan integer (0 hingga 8) menggunakan pustaka Pandas. Penting untuk dicatat bahwa nilai numerik pada Tabel 7 ini bersifat nominal murni. Angka tersebut berfungsi sekadar sebagai identitas pembeda kelas (unique identifier) dan tidak memiliki tingkatan atau hierarki apa pun. Pengodean ke dalam format integer ini merupakan syarat teknis yang mutlak agar data selaras dengan perhitungan metrik loss Sparse Categorical Crossentropy. Selain itu, konversi dari teks ke angka secara drastis memangkas konsumsi memori, yang pada akhirnya mengakselerasi efisiensi komputasi saat mengolah puluhan ribu baris data korpus berita.

Tabel 7. Hasil Label Encoding

Label Kategori	Encoding
news	0
hot	1
finance	2
travel	3
internet	4
health	5

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

otomotif	6
food	7
sport	8

3.1.3 Hasil Tokenisasi dan Padding

Mengingat arsitektur deep learning tidak memiliki kapabilitas untuk memproses untaian teks secara langsung, konversi data leksikal menjadi bilangan bulat (integer) dilakukan melalui mekanisme tokenisasi. Setiap kata diproyeksikan ke dalam sebuah indeks metrik berdasarkan kamus referensi (vocabulary) dari data latih. Distribusi indeks ini diatur oleh tingkat frekuensi; kosakata dengan intensitas kemunculan tertinggi akan dialokasikan pada indeks numerik terkecil. Mengingat variabilitas panjang frasa pada setiap judul berita menghasilkan dimensi sekuens yang heterogen, homogenisasi data mutlak diperlukan. Proses ini dieksekusi melalui teknik padding untuk mematok panjang seluruh urutan menjadi seragam di angka 100 token (max_length). Riset ini secara spesifik mengadopsi metode pre-padding, yakni insersi elemen nol (0) pada awal sekuens yang tidak memenuhi kuota panjang standar. Angka nol tersebut murni diimplementasikan sebagai placeholder (pengisi ruang matriks) dan hampa dari muatan semantik.

Tabel 8. Hasil Tokenisasi & Padding

No	Sebelum Tokenisasi	Setelah Tokenisasi	Setelah Padding
1	1917 menang piala dua untuk best cinematography	[4287, 382, 1699, 208, 11, 2896, 15465]	[0, 0, 0, ..., 4287, 382, 1699, 208, 11, 2896, 15465]
2	mau kerja jadi sales marketing cek gaji di sini	[77, 53, 9, 6673, 9240, 290, 345, 2, 731]	[0, 0, 0, ..., 77, 53, 9, 6673, 9240, 290, 345, 2, 731]
3	sempat kuat ihsg kembali buka merah di 5 425	[150, 359, 360, 121, 36, 583, 2, 18, 12330]	[0, 0, 0, ..., 150, 359, 360, 121, 36, 583, 2, 18, 12330]
4	sensasi paling baru kopi susu campur kecap manis	[2782, 292, 26, 915, 1514, 1515, 4775, 939]	[0, 0, 0, ..., 2782, 292, 26, 915, 1514, 1515, 4775, 939]
5	mu ledek jadi varchester united	[3022, 6674, 9, 15466, 4947]	[0, 0, 0, ..., 3022, 6674, 9, 15466, 4947]

Penyelarasan matriks input merupakan prasyarat fundamental agar jaringan saraf dapat memproses iterasi data secara simultan di setiap batch pelatihan. Tanpa adanya keseragaman dimensi, komputasi kolektif akan gagal dieksekusi. Lebih lanjut, standardisasi ukuran sekuens ini secara signifikan mengakselerasi efisiensi waktu komputasi, sekaligus menjamin kelancaran propagasi informasi dua arah (forward dan backward) yang menjadi karakteristik utama dari model Bi-LSTM. Secara akumulatif, paduan tokenisasi dan padding menyintesis struktur masukan yang ideal dan komprehensif bagi sistem pembelajaran mendalam.

3.1.4 Hasil Pembagian Dataset

Tahap awal pengembangan model diwujudkan melalui partisi korpus data ke dalam dua subset fungsional, yaitu himpunan data latih (training set) dan data uji (testing set). Pemisahan ini merupakan pilar fundamental dalam machine learning untuk mengukur kapabilitas generalisasi sistem terhadap stimulasi informasi baru di luar siklus pembelajaran. Dengan memanfaatkan modul Scikit-Learn, riset ini mengimplementasikan rasio pembagian standar emas sebesar 80:20, sebuah skema proporsional yang dinilai paling ideal dalam menjaga stabilitas antara pemetaan pola semantik multidimensi (9 kelas klasifikasi) dan objektivitas evaluasi akhir. Pengalokasian porsi dominan (80%) bertujuan memberikan ruang komputasi yang luas bagi jaringan saraf dalam menyerap karakteristik unik teks, sedangkan kuota 20% dialokasikan demi menjamin representativitas indikator performa.

Tabel 9. Hasil Pembagian Dataset

No	Jenis Data	Jumlah Data	Persentase (%)
1	Training	72.813	80,0
2	Testing	18.204	20,0

Merujuk pada pemetaan kuantitatif yang didokumentasikan pada Tabel 9, dari agregat data yang menyentuh angka 91.017 sampel, sebanyak 72.813 baris diisolasikan penuh untuk fase optimalisasi bobot arsitektur Bi-LSTM, sementara 18.204 sampel sisanya difungsikan sebagai instrumen verifikasi akurasi. Konfigurasi volume data yang masif ini tidak hanya memperkaya basis pengetahuan model selama fase introduksi pola bahasa, tetapi juga menyediakan landasan empiris yang valid untuk menguji keandalan sistem melalui instrumen pengukuran komprehensif yang melingkupi nilai accuracy, precision, recall, hingga F1-score.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

3.1.5 Hasil Evaluasi Classification Report

Hasil pengujian menggunakan 18.204 data testing menunjukkan bahwa model BI-LSTM yang dikembangkan mampu mencapai tingkat akurasi sebesar 0.85. Capaian tersebut menandakan bahwa model memiliki kemampuan yang baik dalam melakukan klasifikasi terhadap sebagian besar data uji secara tepat. Selain itu, nilai weighted average pada metrik precision, recall, dan f1-score juga tercatat sebesar 0.85. Hasil ini mengindikasikan bahwa performa model relatif konsisten meskipun dataset memiliki distribusi kelas yang tidak merata. Di sisi lain, nilai macro average yang berada pada angka 0.81 menunjukkan masih terdapat perbedaan performa klasifikasi antar kategori. Kondisi tersebut mengindikasikan bahwa beberapa kelas dapat dikenali dengan sangat baik oleh model, sementara beberapa kategori lainnya masih memiliki tingkat klasifikasi yang lebih rendah.

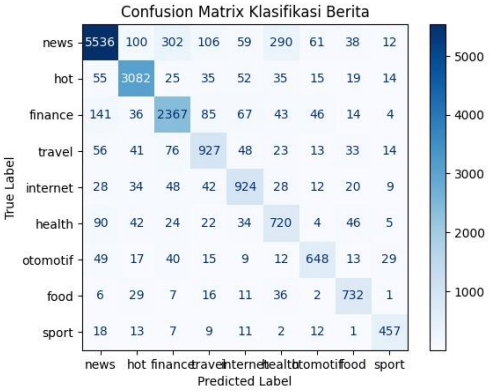
```
=== Classification Report ===
              precision    recall  f1-score   support

        news       0.93      0.85      0.89      6504
         hot       0.91      0.92      0.92      3332
     finance       0.82      0.84      0.83      2803
      travel       0.74      0.75      0.75      1231
    internet       0.76      0.81      0.78      1145
      health       0.61      0.73      0.66       987
    otomotif       0.80      0.78      0.79       832
        food       0.80      0.87      0.83       840
       sport       0.84      0.86      0.85       530

    accuracy                           0.85     18204
   macro avg       0.80      0.82      0.81     18204
weighted avg       0.85      0.85      0.85     18204
```

Gambar 12. Hasil Classification Report

Berdasarkan Gambar 12, kategori hot menunjukkan performa klasifikasi paling optimal dengan nilai f1-score mencapai 0.92. Sebaliknya, kategori health memperoleh performa terendah dengan nilai f1-score sebesar 0.66. Selain itu, kategori news memiliki nilai precision yang relatif tinggi, namun nilai recall yang lebih rendah. Kondisi ini menunjukkan bahwa meskipun prediksi pada kategori tersebut cukup akurat, masih terdapat sejumlah data news yang belum berhasil diidentifikasi dengan baik oleh model.

3.1.6 Hasil Evaluasi Confusion Matrix

Untuk menganalisis pola kesalahan prediksi model, digunakan confusion matrix yang menggambarkan hubungan antara label aktual dan label hasil prediksi. Hasil evaluasi menunjukkan bahwa sebagian besar nilai terkonsentrasi pada diagonal utama, yang menandakan bahwa model mampu mengklasifikasikan data dengan tingkat ketepatan yang cukup baik. Meskipun demikian, masih ditemukan beberapa kesalahan klasifikasi yang cukup menonjol pada kategori tertentu.



Gambar 13. Hasil Confusion Matrix

Berdasarkan Kesalahan klasifikasi paling dominan terjadi pada kategori news yang sering diprediksi sebagai finance dan health, sebagaimana ditunjukkan pada Gambar 13. Selain itu, kategori health juga cukup sering salah diklasifikasikan ke dalam kategori news dan food. Kondisi tersebut menunjukkan adanya kemiripan konteks dan penggunaan istilah antar kategori sehingga model mengalami kesulitan dalam membedakan beberapa kelas tertentu secara tepat.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

3.2 Pembahasan Penelitian

3.2.1 Pembahasan Kinerja Model Bi-LSTM

Hasil evaluasi menunjukkan bahwa model BI-LSTM yang dikembangkan mampu mencapai performa klasifikasi yang cukup baik dengan nilai akurasi sebesar 0.85. Meskipun demikian, performa model pada setiap kategori masih menunjukkan variasi yang cukup signifikan. Perbedaan nilai antara macro average dan weighted average mengindikasikan bahwa model memiliki kecenderungan performa yang lebih optimal pada kategori dengan jumlah data besar, seperti news dan hot, dibandingkan kategori dengan jumlah data lebih sedikit, seperti health dan sport. Kondisi tersebut menunjukkan bahwa distribusi data memberikan pengaruh yang cukup besar terhadap kemampuan model dalam melakukan proses klasifikasi.

3.2.2 Pengaruh Distribusi Data terhadap Performa Model

Perbedaan performa antar kategori memperlihatkan bahwa distribusi jumlah data pada setiap kelas berperan penting dalam proses pembelajaran model. Kategori dengan jumlah data yang lebih besar cenderung menghasilkan performa klasifikasi yang lebih stabil karena model memiliki lebih banyak contoh pola untuk dipelajari. Sebaliknya, kategori dengan jumlah data terbatas cenderung memiliki performa yang lebih rendah akibat keterbatasan variasi data selama proses pelatihan. Hal ini menunjukkan bahwa model lebih mudah mengenali pola pada kelas mayoritas dibandingkan kelas minoritas.

3.2.3 Analisis Kesalahan Klasifikasi pada Kategori News

Apabila dianalisis lebih lanjut, kategori news memiliki nilai precision yang tinggi, namun nilai recall yang relatif lebih rendah. Kondisi ini menunjukkan bahwa model cukup akurat ketika memprediksi kategori news, tetapi masih belum mampu mengenali seluruh data yang sebenarnya termasuk dalam kategori tersebut. Berdasarkan hasil confusion matrix, kesalahan klasifikasi terbesar terjadi ketika data news diprediksi sebagai finance dan health. Fenomena tersebut mengindikasikan adanya kemiripan kosakata maupun topik pembahasan antar kategori sehingga model mengalami kesulitan dalam membedakan konteks secara lebih spesifik.

3.2.4 Tantangan Semantik pada Kategori Health

Kategori health memperoleh performa terendah dengan nilai f1-score sebesar 0.66. Rendahnya nilai precision pada kategori ini menunjukkan bahwa banyak prediksi health sebenarnya berasal dari kategori lain. Selain itu, tingkat kesalahan klasifikasi menuju kategori news dan food juga relatif tinggi. Kondisi tersebut mengindikasikan adanya tumpang tindih semantik antar kategori yang memiliki karakteristik linguistik serupa. Dengan demikian, representasi fitur berbasis tokenisasi yang digunakan masih belum sepenuhnya mampu memahami konteks kalimat secara mendalam.

3.2.5 Konsistensi Performa pada Kategori Hot dan Sport

Berbeda dengan kategori sebelumnya, kategori hot dan sport menunjukkan performa klasifikasi yang relatif stabil dan konsisten. Hal ini dipengaruhi oleh penggunaan kosakata yang lebih khas, spesifik, dan mudah dibedakan dibandingkan kategori lainnya. Karakteristik linguistik yang unik pada kategori tersebut mempermudah model dalam mengenali pola teks sehingga tingkat akurasi klasifikasi menjadi lebih tinggi.

3.2.6 Keterbatasan BI-LSTM dalam Memahami Konteks Bahasa

Meskipun BI-LSTM mampu menangkap dependensi dua arah pada data teks, model ini masih memiliki keterbatasan dalam memahami konteks bahasa yang kompleks. Representasi token yang digunakan masih bersifat sederhana sehingga kemampuan model dalam menangani kalimat ambigu atau kalimat dengan banyak interpretasi makna belum optimal. Akibatnya, kategori yang memiliki tingkat kemiripan semantik tinggi masih sering mengalami kesalahan klasifikasi selama proses prediksi.

4. KESIMPULAN

Kesimpulan penelitian ini membuktikan bahwa arsitektur Bidirectional Long Short-Term Memory (Bi-LSTM) mampu menjadi solusi otomatis yang efektif dalam memetakan sembilan kategori berita berbahasa Indonesia di tengah kondisi ketidakseimbangan data yang ekstrem. Temuan ilmiah utama dari riset ini tidak sebatas pada pencapaian akurasi keseluruhan sebesar 0.85, melainkan pada pembuktian empiris bahwa mekanisme ekstraksi fitur kontekstual dua arah (forward dan backward) terbukti andal dalam mengenali karakteristik kelas dengan kosakata spesifik secara konsisten, seperti pada kategori hot dan sport. Penelitian ini menemukan bahwa disparitas volume sampel yang sangat masif dengan

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

rasio ketimpangan lebih dari tiga belas kali lipat antara kelas mayoritas (news) dan minoritas (sport) berhasil dimitigasi secara terukur. Penerapan class weight sebagai langkah preventif standar pada fase optimasi terbukti mampu memaksa model untuk mendistribusikan atensi pembelajaran secara ekuivalen. Hal ini mengonfirmasi bahwa penanganan bias algoritma pada tahap komputasi sangat krusial untuk mencegah model deep learning mengabaikan kelas minoritas. Meskipun demikian, evaluasi mendalam menyingkap sebuah keterbatasan ilmiah yang penting performa Bi-LSTM mulai mengalami degradasi ketika dihadapkan pada kategori berita yang memiliki tumpang tindih leksikal dan irisan semantik yang tinggi, seperti kebingungan model dalam membedakan teks kategori health dengan news atau food. Kontribusi penelitian ini diharapkan dapat menjadi landasan evaluatif bagi pengembangan sistem NLP berbahasa Indonesia, sekaligus merekomendasikan perlunya integrasi metode embedding tingkat lanjut untuk mengatasi anomali kemiripan semantik pada riset di masa mendatang.

UCAPAN TERIMAKASIH

Ucapan terima kasih penulis sampaikan kepada Dosen Pembimbing atas bimbingan, arahan, dan dukungan yang diberikan selama proses penyusunan artikel ini sehingga penelitian dapat diselesaikan dengan baik.

REFERENCES

[1] H. Mustofa and A. A. Mahfudh, “Klasifikasi Berita Hoax Dengan Menggunakan Metode Naive Bayes,” *Walisongo Journal of Information Technology*, vol. 1, no. 1, p. 1, Nov. 2019, doi: 10.21580/wjit.2019.1.1.3915.

[2] B. P. Nayoga, R. Adipradana, R. Suryadi, and D. Suhartono, “Hoax Analyzer for Indonesian News Using Deep Learning Models,” *Procedia Comput. Sci.*, vol. 179, pp. 704–712, 2021, doi: 10.1016/j.procs.2021.01.059.

[3] D. R. Alghifari, M. Edi, and L. Firmansyah, “Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia,” *Jurnal Manajemen Informatika (JAMIKA)*, vol. 12, no. 2, pp. 89–99, Sep. 2022, doi: 10.34010/jamika.v12i2.7764.

[4] A. H. Pradhana, E. Daniati, and M. N. Muzaki, “Penerapan Bi-LSTM Untuk Named Entity Recognition Pada Teks Bahasa Indonesia,” *The Indonesian Journal of Computer Science Research*, vol. 4, no. 2, pp. 96–106, 2025.

[5] Anas Fikri Hanif, Theopilus Bayu Sasongko, and Arif Dwi Laksito, “Perbandingan Kinerja LSTM, Bi-LSTM, dan GRU pada Klasifikasi Judul Berita Clickbait,” *The Indonesian Journal of Computer Science*, vol. 12, no. 4, Aug. 2023, doi: 10.33022/ijcs.v12i4.3281.

[6] C. N. Daiman, A. Yuniar Rahman, and F. Nudiyansyah, “KLASIFIKASI TEKS BERITA BREAKING NEWS DI MANGGARAI MENGGUNAKAN LONG SHORT TERM MEMORY (LSTM),” *Jurnal Mnemonic*, vol. 7, no. 2, pp. 170–174, Jun. 2024, doi: 10.36040/mnemonic.v7i2.9939.

[7] T. Young, D. Hazarika, S. Poria, and E. Cambria, “Recent Trends in Deep Learning Based Natural Language Processing [Review Article],” *IEEE Comput. Intell. Mag.*, vol. 13, no. 3, pp. 55–75, Aug. 2018, doi: 10.1109/MCI.2018.2840738.

[8] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, “Text Classification Algorithms: A Survey,” *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.

[9] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

[10] S. Wang, L. L. Minku, and X. Yao, “Resampling-Based Ensemble Methods for Online Class Imbalance Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 5, pp. 1356–1368, May 2015, doi: 10.1109/TKDE.2014.2345380.

[11] D. W. Otter, J. R. Medina, and J. K. Kalita, “A Survey of the Usages of Deep Learning for Natural Language Processing,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 604–624, Feb. 2021, doi: 10.1109/TNNLS.2020.2979670.

[12] C. Huang, Y. Li, C. C. Loy, and X. Tang, “Learning Deep Representation for Imbalanced Classification,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2016, pp. 5375–5384. doi: 10.1109/CVPR.2016.580.

[13] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.

[14] Dianda Rifaldi *et al.*, “Evaluasi Sentimen Pengguna ChatGPT Menggunakan Naive Bayes: Tinjauan dari Confusion Matrix dan Classification Report,” *Jurnal Riset Sistem dan Teknologi Informasi*, vol. 3, no. 2, pp. 81–89, Jul. 2025, doi: 10.30787/restia.v3i2.1990.

[15] H. Oktavianto, H. W. Sulistyo, G. Wijaya, D. Irawan, and G. Abdurrahman, “Analisis Perbandingan Decision Tree dan Random Forest Pada Klasifikasi Teks Data Kesehatan,” *BINA INSANI ICT JOURNAL*, vol. 11, no. 1, p. 56, Jun. 2024, doi: 10.51211/biict.v11i1.2928.

[16] K. D. Ningtyas, R. Kurniawan, and A. Armansyah, “Penerapan Natural Language Processing Pada Aplikasi Chatbot Info Layanan Kantor Menggunakan Naive Bayes Algorithm,” *J-SISKO TECH (Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD)*, vol. 6, no. 1, p. 266, Jan. 2023, doi: 10.53513/jsk.v6i1.7413.

[17] H. Eldo, A. Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, “Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online,” *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, Oct. 2024, doi: 10.33395/jmp.v13i2.14186.

[18] A. K. Uysal and S. Gunal, “The impact of preprocessing on text classification,” *Inf. Process. Manag.*, vol. 50, no. 1, pp. 104–112, Jan. 2014, doi: 10.1016/j.ipm.2013.08.006.

[19] M. Y. Dhinora and E. Mailoa, “Analisa Tweet Mahasiswa untuk Deteksi Gejala Depresi dengan Penerapan Natural Language Processing,” *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 6, no. 2, pp. 1193–1211, May 2025, doi: 10.63447/jimik.v6i2.1405.

Management of Information System Journal
Vol 4, No 3, Juli 2026, Hal. 127 - 140
ISSN: 2964-9455 (media online)
DOI: 10.47065/mis.v2i3.2745
https://journal.fkpt.org/index.php/MIS

[20] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.
[21] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a Few Examples," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021, doi: 10.1145/3386252.
[22] L. Qin *et al.*, "Large language models meet NLP: a survey," *Front. Comput. Sci.*, vol. 20, no. 11, p. 2011361, Nov. 2026, doi: 10.1007/s11704-025-50472-3.
[23] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text Classification Algorithms: A Survey," *Information*, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.
[24] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Sep. 2013.
[25] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Sci. China Technol. Sci.*, vol. 63, no. 10, pp. 1872–1897, Oct. 2020, doi: 10.1007/s11431-020-1647-3.
[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[27] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.
[28] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning--based Text Classification," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–40, Apr. 2022, doi: 10.1145/3439726.
[29] Y. Lou, J. Zhou, J. Zhou, D. Ji, and Q. Zhang, "Emoji multimodal microblog sentiment analysis based on mutual attention mechanism," *Sci. Rep.*, vol. 14, no. 1, p. 29314, Nov. 2024, doi: 10.1038/s41598-024-80167-x.
[30] K. Hara, D. Saito, and H. Shouno, "Analysis of function of rectified linear unit used in deep learning," in *2015 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Jul. 2015, pp. 1–8. doi: 10.1109/IJCNN.2015.7280578.
[31] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't Decay the Learning Rate, Increase the Batch Size," Feb. 2018.
[32] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Jan. 2017.
[33] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019, doi: 10.1186/s40537-019-0192-5.