



## Klasifikasi teks clickbait dengan metode deep learning hybrid lstm-cnn

Aditya Darma<sup>1</sup>, Sepni<sup>2</sup>, Shelly Adelia Sary<sup>3</sup>, Yayang Uswatun Hasanah<sup>4</sup>

Email: <sup>1</sup>210401174@student.umri.ac.id, <sup>2</sup>210401168@student.umri.ac.id, <sup>3</sup>210401164@student.umri.ac.id,

<sup>4</sup>210401169@student.umri.ac.id

<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau

Diterima: 05 April 2020 | Direvisi: 05 Mei 2020 | Disetujui: 27 Mei 2020

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,  
Universitas Muhammadiyah Riau, Indonesia

### Abstrak

Penelitian ini bertujuan untuk melakukan Penentuan kategori judul berita dilakukan dengan membaginya ke dalam dua kelompok, yaitu clickbait dan non-clickbait menggunakan metode *hybrid LSTM-CNN*. Dataset yang digunakan terdiri dari 14.878 data dalam dua kategori dengan berita *clickbait* 6.285 data dan berita non clickbait 8.693 data yang diperoleh dari laman kaggle. Tahapan penelitian meliputi preprocessing data melalui proses *cleaning*, *tokenizing*, *stopword removal*, *stemming*, dan representasi teks menggunakan algoritma *Word2Vec*. Dataset kemudian akan di pisahkan kedalam bentuk data latih dan uji menggunakan rasio 80:20. Model *hybrid LSTM-CNN* digunakan karena keunggulan CNN dalam mengekstraksi fitur lokal serta kemampuan LSTM dalam memahami hubungan sekuensial antar kata. Evaluasi kinerja model dilakukan menggunakan confusion matrix, dengan hasil akurasi sebesar 77,07%, recall 70%, presisi 73%, dan F1-score 71%. Model *hybrid LSTM-CNN* menunjukkan performa yang lebih baik dibandingkan model terpisah dengan peningkatan akurasi dari 77% menjadi 77,07%. Penelitian ini menunjukkan bahwa kombinasi model LSTM-CNN efektif dalam menangani klasifikasi teks berita clickbait dan non-clickbait, memberikan hasil yang cukup baik dalam meningkatkan performa model sebelumnya.

**Kata kunci:** *LSTM-CNN, klasifikasi teks, judul berita clickbait dan non clickbait, deep learning*

### *Clickbait text classification with deep learning hybrid lstm-cnn method*

#### *Abstract*

*This study aims to determine the category of news titles by dividing them into two groups, namely clickbait and non-clickbait using the LSTM-CNN hybrid method. The dataset used consists of 14,878 data in two categories with 6,285 clickbait news data and 8,693 non clickbait news data obtained from the kaggle page. The research stages include data preprocessing through cleaning, tokenizing, stopwords removal, stemming, and text representation using the Word2Vec algorithm. The dataset will then be separated into training and test data using a ratio of 80:20. The LSTM-CNN hybrid model is used because of CNN's advantage in extracting local features as well as LSTM's ability to understand sequential relationships between words. The model performance evaluation was conducted using confusion matrix, with the results of 77.07% accuracy, 70% recall, 73% precision, and 71% F1-score. The LSTM-CNN hybrid model showed better performance than the separate models with an increase in accuracy from 77% to 77.07%. This research shows that the LSTM-CNN model combination is effective in handling clickbait and non-clickbait news text classification, providing quite good results in improving the performance of the previous model.*

**Keywords:** *LSTM-CNN, text classification, clickbait and non clickbait on news headlines, deep learning*

## 1. PENDAHULUAN

Seiring perkembangan zaman peningkatan penggunaan internet semakin meningkat dilingkungan masyarakat. Hal ini menyebabkan berita online kini menjadi pilihan utama dalam mengakses informasi. Berita online sangat mudah diakses dan disebarluaskan melalui berbagai perangkat, namun kemudahan tersebut sering kali dimanfaatkan oleh sejumlah produsen berita untuk menyesatkan para pembaca [1]. Fenomena menyesatkan ini terlihat dari ketidaksesuaian antara informasi yang disampaikan dengan judul artikel atau berita di media elektronik. Judul yang dapat dikatakan menjebak, atau dikenal dengan istilah *clickbait*, digunakan untuk menarik perhatian pembaca demi meningkatkan jumlah kunjungan ke situs. Semakin banyak pengunjung, semakin besar pula potensi keuntungan yang didapatkan, terutama dari pendapatan iklan [2].

Salah satu faktor utama yang mendorong maraknya penggunaan *clickbait* di media online adalah tingginya persaingan antar media untuk memperoleh pembaca [3]. Judul pada berita *clickbait* sangat sering memberikan gambaran keliru, baik melalui judul maupun gambar yang digunakan sehingga menciptakan ekspektasi yang tidak sesuai dengan isi berita. Ketika pembaca mengakses laman tersebut mereka sering merasa kecewa karena isi berita tidak memenuhi harapan yang diciptakan oleh judulnya.

Berdasarkan permasalahan tersebut upaya pencegahan terhadap berita *clickbait* menjadi penting. Pendekatan yang bisa diterapkan adalah dengan mengelompokkan data melalui proses klasifikasi terhadap berita untuk menentukan apakah berita tersebut termasuk kategori *clickbait* atau *non clickbait*. Pendekatan ini dapat membantu pembaca menyeleksi konten yang beredar di media elektronik dengan lebih cepat. Pada penelitian terdahulu, algoritma *Long Short-Term Memory* (LSTM) telah digunakan untuk klasifikasi berita *clickbait* dan *non-clickbait* dengan menghasilkan akurasi sebesar 77%. Namun, untuk meningkatkan akurasi model tersebut, penelitian ini berfokus pada pengoptimalan LSTM dengan mengkombinasikannya bersama *Convolutional Neural Network* (CNN). CNN memiliki keunggulan dalam ekstraksi fitur yang lebih efektif, sementara LSTM baik digunakan dalam memahami hubungan yang terdapat didalam sebuah teks dokumen. Kombinasi kedua arsitektur ini, yang dikenal sebagai *Convolutional Neural Network-Long Short-Term Memory* (CNN-LSTM) telah terbukti memberikan hasil yang lebih baik dibandingkan penggunaan CNN atau LSTM secara terpisah [4].

Penelitian sebelumnya juga menunjukkan bahwa metode *hybrid* LSTM-CNN memiliki potensi besar dalam klasifikasi teks. Sebagai contoh, penggunaan metode ini dalam pengklasifikasian komentar di aplikasi X terkait sentimen publik Indonesia terhadap konflik Palestina-Israel mencapai akurasi sebesar 96% [5]. Selain itu, penelitian lainnya membuktikan bahwa metode *hybrid* LSTM-CNN mampu meningkatkan akurasi sebesar 0,99% dibandingkan metode standar dalam klasifikasi konten media berita online di Indonesia. Dengan hasil-hasil tersebut, metode *hybrid* LSTM-CNN dinilai cocok untuk meningkatkan akurasi model dalam klasifikasi berita *clickbait* dan *non-clickbait* [6].

Secara umum, *clickbait* diartikan sebagai judul berita yang dirancang untuk menarik perhatian dan membuat pembaca tertarik mengklik tautan menuju situs utama [7]. Fenomena *clickbait* ini dinilai merugikan praktik jurnalistik, karena jurnisme yang seharusnya mengutamakan kepentingan publik justru sering mengabaikan nilai berita dan verifikasi demi mengejar jumlah klik atau *viewers* untuk keuntungan korporasi [8]. Oleh karena itu, optimalisasi metode *hybrid* LSTM-CNN menjadi langkah strategis untuk meningkatkan performa model klasifikasi teks. Dalam penelitian mengenai penggabungan antara CNN dan LSTM untuk menganalisa performanya pada konteks media berita online di Indonesia, metode ini terbukti mampu meningkatkan akurasi model sebesar 0,99% dibandingkan penggunaan metode standar. Kombinasi CNN dan LSTM memanfaatkan keunggulan CNN dalam mengenali fitur lokal pada teks, sekaligus memanfaatkan kemampuan LSTM untuk memahami konteks narasi yang panjang.

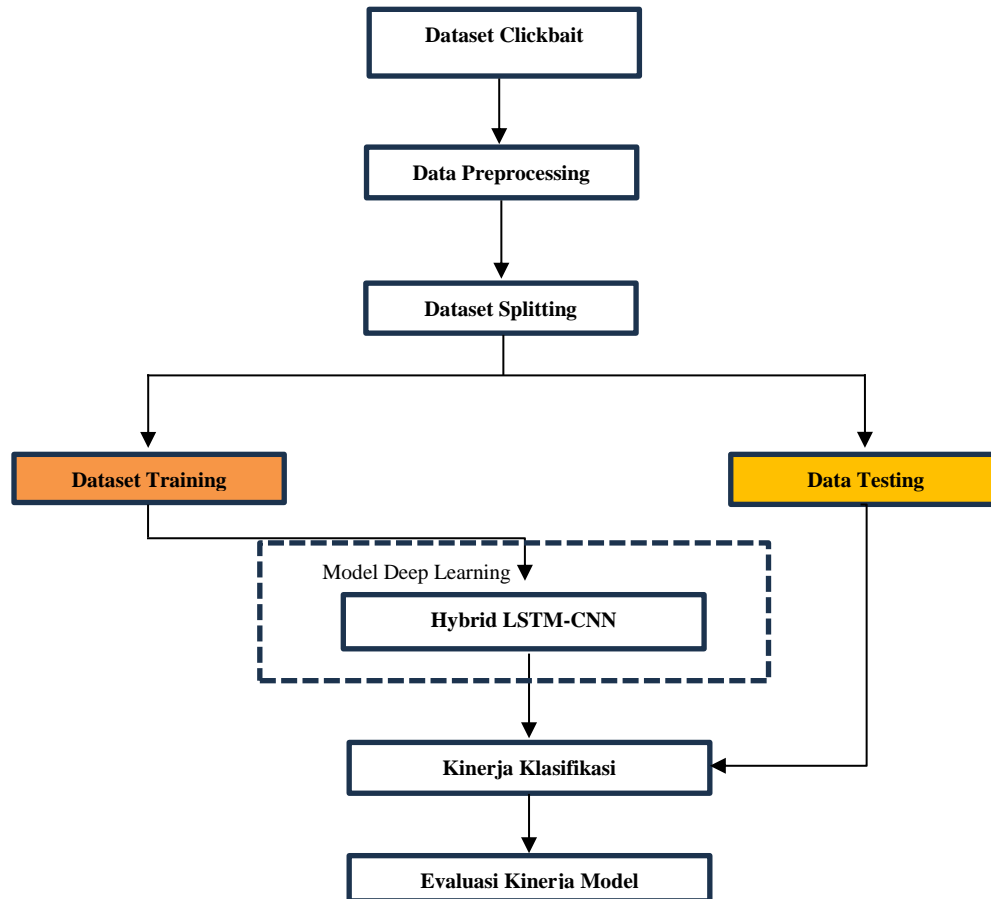
Penelitian ini bertujuan untuk menjawab tantangan dalam mengoptimalkan performa LSTM dengan menggunakan metode *hybrid* LSTM-CNN. Diharapkan metode ini dapat meningkatkan kinerja model dalam klasifikasi judul berita *clickbait* dan *non-clickbait*. Dengan demikian, hal ini mampu memudahkan masyarakat dalam membedakan kategori berita tersebut, yang pada akhirnya dapat mengurangi kesalahan persepsi dari berita yang beredar di dunia maya.

Penelitian ini dilakukan untuk dapat menjawab pertanyaan berikut :

1. Apa langkah-langkah yang diperlukan untuk mengoptimalkan model Long Short-Term Memory (LSTM) dalam klasifikasi berita *clickbait* dan *non-clickbait* menggunakan pendekatan hybrid dengan Convolutional Neural Network (CNN)?
2. Apakah optimasi menggunakan metode *hybrid* Long Short-Term Memory (LSTM) dan Convolutional Neural Network (CNN) dapat meningkatkan akurasi model LSTM dalam mengidentifikasi judul berita *clickbait* dibandingkan dengan model LSTM yang tidak dioptimasi?

## 2. METODE PENELITIAN

Penelitian ini dirancang untuk melakukan optimasi performa model *Long Short Term Memory* (LSTM) dengan menggunakan metode *hybrid Long Short Term-Memory* (LSTM) dan *Convolutional Neural Network* (CNN) untuk klasifikasi judul berita *clickbait* dan *non clickbait*. metodologi penelitian melibatkan beberapa tahapan utama yang ditunjukkan pada Gambar 1 berikut ini :



Gambar 1. Tahapan Penelitian

### 2.1. Dataset Clickbait

Dataset yang digunakan dalam studi kasus ini berasal dari kaggle. Dataset berisi 14.1978 data judul berita dengan kategori *clickbait* dan *non clickbait* dan memiliki 2 fitur. Tabel 1 berikut akan menampilkan informasi fitur dari dataset yang digunakan.

Tabel 1. Struktur Dataset

No	Fitur	Tipe Data	Keterangan
1	Title	String	Judul Berita yang ditampilkan
2	Label	String	Keterangan Jenis Judul Berita ( <i>Clickbait</i> dan <i>Non Clickbait</i> )

### 2.2. Preprocessing

*Preprocessing* adalah langkah penting dalam pengolahan dataset *clickbait* untuk memastikan data siap digunakan oleh model *deep learning*. Pemrosesan data akan mengolah data mentah yang diperoleh dari suatu sumber menjadi informasi yang lebih terstruktur dan siap digunakan untuk analisis lanjutan [9]. Beberapa langkah yang dilakukan dalam tahap *preprocessing* ini adalah sebagai berikut :

### 2.2.1 Cleaning

Pada tahap ini akan dilakukan proses pembersihan dan perbaikan pada suatu kumpulan data yang digunakan. proses ini meliputi *case folding* dengan mengubah karakter pada sebuah dokumen menjadi huruf kecil dengan tujuan untuk mengurangi variasi kata yang ditemukan pada sebuah dataset. selain itu, terdapat pula beberapa tahap lain yang dilakukan seperti pembersihan simbol, *URL*, tanda baca, dan beberapa elemen lainnya tidak relevan untuk digunakan pada tahap analisa.

### 2.2.2 Tokenizing

Pada tahap tokenisasi, di mana sebuah kalimat akan dibagi menjadi unit-unit lebih kecil yang disebut token. Langkah ini bertujuan untuk menganalisis kata-kata yang terdapat dalam sebuah kalimat [10]. Beberapa jenis *tokenizing* yang cukup umum digunakan adalah pemotongan kata, pemotongan frase serta pemotongan karakter.

### 2.2.3 Stopword

*Stopwords* adalah kata-kata yang tidak memiliki keunikan dan mencerminkan karakteristik dari data yang sedang diproses. Dalam tahap ini, kata-kata yang tidak diperlukan akan dihapus dari data. Selain menggunakan kamus *stopwords* yang sudah ada, kita juga dapat menambahkan kamus kata yang dibuat secara pribadi [11].

### 2.2.4 Stemming

Stemming merupakan proses mengubah token berimbuhan menjadi bentuk dasar kata dengan menghapus semua imbuhan yang melekat pada token tersebut. Proses stemming sangat penting untuk menghapus imbuhan yang terletak di awal atau di akhir kata.

### 2.2.5 Word2Vec

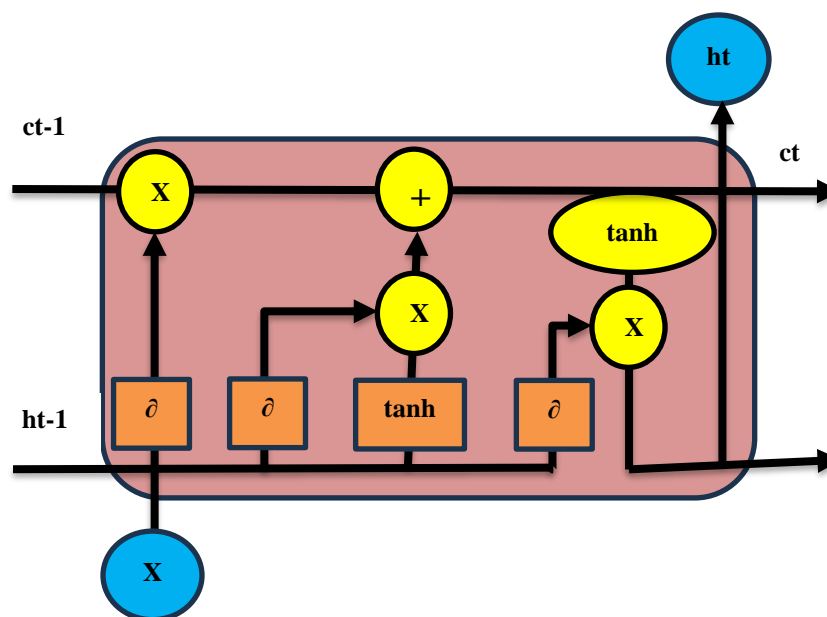
*Word2Vec* adalah salah satu metode embedding yang diterapkan dalam proses pra-pelatihan untuk klasifikasi data. Metode ini memiliki beberapa jenis arsitektur seperti *Skip-gram* dan *Continuous Bag of Words (CBOW)*. Setiap arsitektur dari *Word2Vec* mempunyai peranan yang penting dalam menghasilkan performa model yang baik dalam metode *deep learning* [12]. Sebuah penelitian mengenai perbandingan antara metode Tf-Idf dan *Word2Vec* dalam klasifikasi teks sentimen masyarakat memberikan hasil bahwa algoritma dengan penggunaan *Word2Vec* sebagai representasi teksnya menghasilkan hasil yang lebih baik dibandingkan dengan menggunakan metode Tf-Idf. algoritma yang dikombinasikan dengan *Word2Vec* mendapatkan nilai F1 - Score lebih baik yaitu sebesar 0.941% dibandingkan algoritma yang dikombinasikan dengan Tf-Idf yang mendapatkan hasil F1 - Score sebesar 0.940% [13]. Sehingga dapat dinilai bahwa penggunaan *Word2Vec* lebih unggul untuk digunakan dalam proses representasi teks pada model algoritma kecerdasan buatan yang akan digunakan.

## 2.3 Data Splitting

Data splitting adalah tahap di mana data dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian. Dalam penelitian ini, pembagian data dilakukan dengan rasio 80% untuk data pelatihan dan 20% untuk data pengujian. Data pelatihan digunakan untuk membangun model, sedangkan data pengujian dipakai untuk mengevaluasi kinerja model yang telah dibuat [14].

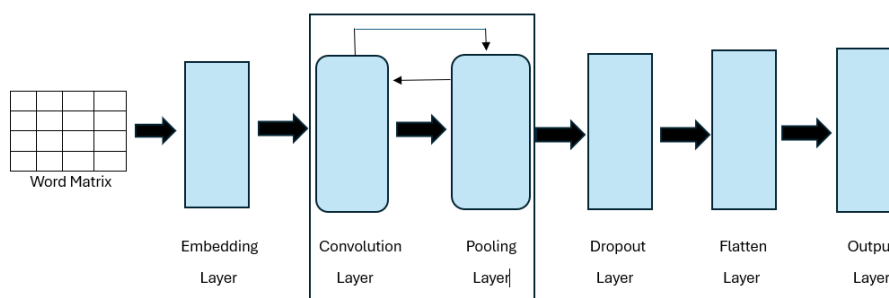
## 2.4 Model LSTM dan CNN

*LSTM* merupakan algoritma pengembangan dari *Recurrent Neural Network (RNN)*[15]. Model *LSTM* memiliki struktur yang rumit dan unit memori yang dapat diatur. Pada penggunaannya algoritma *LSTM* dapat mengingat informasi dengan baik dengan rentang waktu yang lama [16] . Pada desain *LSTM* terdapat beberapa komponen penting seperti *input gate*, *forget gate*, dan *output gate*. *LSTM* memiliki struktur yang kompleks yang mendukung penyimpanan dan akses data jangka panjang, serta kemampuan untuk menangani masalah yang bersifat *fuzzy*. pada penelitian ini, fitur Tf-Idf digunakan sebagai masukan untuk model *LSTM* yang kemudian akan memberikan hasil akurasi dari proses klasifikasi teks judul berita *clickbait*.



Gambar 2. Arsitektur Model LSTM

CNN adalah Algoritma *Deep Learning* yang sangat efektif dalam analisis gambar. Algoritma ini memanfaatkan lapisan konvolusi yang secara otomatis dan adaptif belajar untuk mengidentifikasi fitur-fitur penting dari sebuah gambar untuk tugas seperti klasifikasi dan deteksi objek.



Gambar 3. Arsitektur Model CNN

Pada Gambar 3 Arsitektur Model CNN terdiri dari tiga kategori lapisan yaitu: lapisan Input, lapisan tersembunyi, dan lapisan output. Lapisan tersembunyi dalam jaringan saraf mencakup lapisan konvolusi bergerak melintasi matriks input untuk menghasilkan peta fitur yang akan digunakan sebagai input untuk lapisan selanjutnya. Kemudian, ditambahkan lapisan-lapisan lain seperti lapisan pooling, lapisan yang sepenuhnya terhubung dan lapisan normalisasi.

Metode CNN umumnya digunakan untuk analisis citra atau gambar karena lapisan konvolusinya efektif dalam mengekstraksi fitur-fitur kecil dari sebuah gambar. Meskipun demikian, sejumlah studi telah mengadaptasi CNN untuk pengolahan teks berita. Berbagai penelitian menunjukkan bahwa CNN memiliki kinerja yang lebih unggul dibandingkan metode lainnya dalam pengolahan teks. Diketahui bahwa lapisan konvolusi pada CNN dapat mengurangi beban komputasi pada jaringan saraf.

## 2.5 Evaluasi Kinerja Model

Untuk mengetahui sebuah model klasifikasi telah tergolong memiliki performa yang baik atau tidak dapat dilakukan evaluasi menggunakan *confusion matrix*. Confusion matrix merupakan alat untuk membandingkan hasil prediksi model dengan label yang benar, dan matriks ini menggambarkan sejauh mana akurasi model dalam mengklasifikasikan data serta banyaknya kesalahan yang terjadi. Berikut persamaan yang digunakan dalam mengevaluasi model:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1\ Score = 2x \frac{\Sigma Recall \times \Sigma Precision}{\Sigma Recall + \Sigma Precision}$$

Dalam evaluasi performa model LSTM - CNN digunakan algoritma evaluasi confusion matrix, yang terbagi atas kategori hasil prediksi yaitu *True Positive* (TP), *True Negative* (TN), *False Negative* (FN), dan *False Positive* (FP).

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Pengumpulan dataset

Dalam penelitian ini, data yang digunakan diperoleh dari situs Kaggle yang dapat diakses melalui tautan berikut.[\[Link\]](#). Dataset tersebut berisi 14.878 data dengan label *clickbait* sebanyak 6285 dan label *non clickbait* sebanyak 8693. tabel 2 berikut akan menampilkan dari dataset yang digunakan pada penelitian.

Tabel 2. Dataset Clickbait Dan Non Clickbait

title	label	label_score
Masuk Radar Pilwalkot Medan, Menantu Jokowi Bertemu DPW NasDem Sumut	non clickbait	0
Malaysia Sudutkan RI: Isu Kabut Asap hingga Invasi Babi	non clickbait	0
Viral! Driver Ojol di Bekasi Antar Pesanan Makanan Pakai Sepeda	clickbait	1
Kemensos Salurkan Rp 7,3 M bagi Korban Kerusakan Sosial di Papua	non - clickbait	0
Terkait Mayat Bayi Mengenakan di Tangerang, Seorang Pria Ditangkap Polisi	non - clickbait	0
Uang Rp 1,6 M yang Raib di Parkiran Kantor Gubernur Sumut untuk Honor TAPD	non - clickbait	0

#### 3.2. Preprocessing

##### - Cleaning

pada tahapan ini akan dilakukan pembersihan dataset yang akan digunakan. proses pembersihan data meliputi *case folding*, *URL*, tanda baca, simbol dan hal lain yang tidak relevan digunakan pada tahap analisa. tabel 3 berikut ini akan menampilkan contoh data sebelum dan sesudah dilakukan pembersihan.

Tabel 3. Dataset Sebelum Dan Setelah Proses Cleaning

Sebelum cleaning	Setelah cleaning
Masuk Radar Pilwalkot Medan, Menantu Jokowi Bertemu DPW NasDem Sumut	masuk radar pilwalkot medan menantu jokowi bertemu dpw nasdem sumut

Malaysia Sudutkan RI: Isu Kabut Asap hingga Invasi Babi	malaysia sudutkan ri isu kabut asap hingga invasi babi
Viral! Driver Ojol di Bekasi Antar Pesanan Makanan Pakai Sepeda	viral driver ojol di bekasi antar pesanan makanan pakai sepeda
Kemensos Salurkan Rp 7,3 M bagi Korban Kerusuhan Sosial di Papua	kemensos salurkan rp 7 3 m bagi korban kerusuhan sosial di papua
Terkait Mayat Bayi Mengenaskan di Tangerang, Seorang Pria Ditangkap Polisi	terkait mayat bayi mengenaskan di tangerang seorang pria ditangkap polisi

- *Tokenizing*

Pada tahapan ini akan dilakukan pemecahan kalimat yang ada pada dataset menjadi unit yang lebih kecil dengan tujuan untuk mengeksplor kata - kata yang terdapat pada dataset yang digunakan. berikut ini adalah gambaran dari dataset yang telah dilakukan proses tokenisasi.

Tabel 4. Dataset Setelah Proses Tokenizing

Sesudah Tokenezing
masuk radar pilwalkot medan menantu jokowi bertemu dpw nasdem sumut
malaysia sudutkan ri isu kabut asap hingga invasi babi
viral driver ojol di bekasi antar pesanan makanan pakai sepeda
kemensos salurkan rp 7 3 m bagi korban kerusuhan sosial di papua
terkait mayat bayi mengenaskan di tangerang seorang pria ditangkap polisi

- *Stopword dan Stemming*

Pada tahap ini, akan dilakukan penyaringan terhadap kata-kata yang tidak mencerminkan karakteristik data menggunakan perintah stopwords serta pengembalian kata ke bentuk dasarnya melalui proses stemming. Kedua langkah ini bertujuan untuk memastikan bahwa hanya kata-kata penting yang digunakan dan mengurangi variasi kata dalam data. Gambaran mengenai dataset yang telah melalui proses penghapusan stopwords dan stemming dapat dilihat pada Tabel 5 berikut ini.

Tabel 5. Dataset setelah proses *Stopword* dan *Stemming*

Clean text
Masuk radar pilwalkot medan menantu jokowi temu dpw nasdem sumut
malaysia sudut ri isu kabut asap invasi babi
viral driver ojol bekas antar pesan makan pakai sepeda
kemensos salur rp 7 3 m korban rusuh sosial papua



kait mayat bayi enas tangerang orang pria tangkap polisi
tolak ruu tanah ribu tani siap gelar aksi istana dpr esok
ada niat tambah momong tanpa ikut program hamil fedri nuril kalau penguin ya buat aja
edar isu internet papua blokir lagi telkom minta warga tak percaya hoaks
txt akan segera comeback soobin aku gatal ingin pamer spoiler
emmy awards 2019 peter dinklage cetak rekor baru berkat peran tyron lannister game of thrones

#### - Word2Vec

pada tahapan ini akan dilakukan representasi dari teks yang sebelumnya telah dilakukan pembersihan dari nilai yang tidak relevan digunakan untuk proses analisa. algoritma ini akan merepresentasikan teks dengan cara mengelompokkan kata yang serupa yang memiliki vektor yang sama dengan menggunakan neural network.

### 3.3. Data Splitting

Data Splitting merupakan langkah untuk membagi dataset menjadi beberapa bagian, yaitu untuk pelatihan (training), validasi (validation), dan pengujian (testing) dalam machine learning atau analisis statistik. Proses ini sangat penting untuk memastikan bahwa model yang dikembangkan dapat bekerja dengan baik saat menghadapi data baru yang belum pernah diproses sebelumnya. Pembagian data dengan rasio yang umum diterapkan adalah 80:20, di mana 80% digunakan untuk pelatihan dan 20% untuk pengujian. Selain itu, pembagian dengan rasio 70:30, 60:40, 90:10, bahkan 50:50 juga sering digunakan. [optimal ratio for data splitting].

### 3.4 Model Hybrid LSTM-CNN

Model Hybrid LSTM-CNN dikombinasikan karena kedua Algoritma dinilai berpotensi dapat menyelesaikan permasalahan pengklasifikasian teks karena menggabungkan kemampuan CNN dalam mengidentifikasi fitur lokal dalam teks serta memanfaatkan kemampuan LSTM dalam pemahaman sekuensi narasi yang lebih panjang.

CNN memiliki keunggulan dalam proses ekstraksi fitur yang lebih efektif. Disisi lain, LSTM ini dipilih dikarenakan kemampuannya untuk mengidentifikasi keterkaitan antar konteks pada dokumen teks. menggabungkan dua arsitektur *deep learning*, yaitu CNN dan LSTM, untuk klasifikasi teks berita. Kombinasi arsitektur ini dinamakan *Convolutional Long Short-Term Memory* (C-LSTM). Menunjukkan bahwa akurasi C-LSTM melebihi akurasi yang diperoleh dari CNN dan LSTM ketika digunakan secara terpisah. Berikut adalah arsitektur dari model LSTM-CNN yang digunakan pada penelitian ini

Tabel 6. Arsitektur model LSTM - CNN

Define INPUT_LAYER with shape (max_sequence_length)
Create EMBEDDING_LAYER:
Input dimension: len(word_index) + 1
Output dimension: embedding_dim
Weights: embedding_matrix
Trainable: False
Build LSTM_BRANCH:
Add LSTM_LAYER with 128 units and return_sequences=True
Apply GLOBAL_MAX_POOLING_1D to LSTM output
Build CNN_BRANCH:



```

Add CONV1D_LAYER with 64 filters, kernel_size=3, activation='relu'
Add MAX_POOLING1D with pool_size=2
Flatten CNN output

CONCATENATE LSTM_BRANCH and CNN_BRANCH outputs

Add DENSE_LAYER:
    Units: 64
    Activation: ReLU
    Regularization: L2 (0.001)

Add DROPOUT_LAYER with rate=0.5

Add OUTPUT_LAYER:
    Units: 1
    Activation: Sigmoid

Build MODEL:
    Inputs: INPUT_LAYER
    Outputs: OUTPUT_LAYER

COMPILE MODEL:
    Optimizer: Adam
    Loss: Binary Crossentropy
    Metrics: Accuracy

Print MODEL SUMMARY
    
```

### 3.5 Evaluasi Kinerja Model

Evaluasi kinerja model diperlukan untuk mengetahui apakah model yang dibangun sudah baik atau tidak. Pada penelitian ini evaluasi model pengklasifikasian teks judul berita *clickbait* dan *non clickbait* menggunakan *confusion matriks* yang menghasilkan nilai akurasi 77.07 %, recall 70%, presisi 73%, dan f1 score 71%

Tabel 7. Confusion matrix

		Nilai Prediksi	
		Non Clickbait	Clickbait
Nilai Aktual	Not Clickbait	1449	314
	Clickbait	373	860

#### **4. KESIMPULAN**

Dalam penelitian ini, penggunaan metode *hybrid* LSTM-CNN dalam kasus pengklasifikasian teks pada dataset judul berita *clickbait* dan *non clickbait*, dinilai sudah baik dalam meningkatkan performa model. Hasil evaluasi pada model dengan metode *hybrid* LSTM-CNN ini memperoleh nilai akurasi sebesar 77.07 %, sedangkan hasil evaluasi sebelum penerapan metode *hybrid* LSTM-CNN memperoleh nilai akurasi sebesar 77%. Penelitian ini menunjukkan bahwa kemampuan model dengan penerapan metode *hybrid* LSTM-CNN dalam mengklasifikasikan judul berita *clickbait* dan *non clickbait* mampu meningkatkan performa model sebelumnya dan mencapai tingkat keberhasilan yang cukup baik.

#### **Ucapan Terimakasih**

Penulis menyampaikan rasa terima kasih yang mendalam kepada Fakultas Ilmu Komputer, Universitas Muhammadiyah Riau, atas dukungan fasilitas dan suasana akademik yang mendukung selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada seluruh dosen, tenaga kependidikan, serta rekan-rekan mahasiswa di Fakultas Ilmu Komputer yang telah memberikan dorongan serta bantuan yang sangat berarti sepanjang proses penelitian. Tanpa kontribusi dan dukungan dari berbagai pihak yang terlibat, tentu saja pelaksanaan penelitian dan penulisan jurnal ini tidak akan dapat terlaksana dengan baik

**DAFTAR PUSTAKA**

- [1] W. Afandi, S. N. Saputro, A. M. Kusumaningrum, H. Ardiansyah, M. H. Kafabi, and S. Sudianto, "Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM," vol. 7, no. 2, 2022.
- [2] R. Sagita, U. Enri, and A. Primajaya, "Klasifikasi Berita Clickbait Menggunakan K-Nearest Neighbor (KNN)," *JOINS (Journal of Information System)*, vol. 5, no. 2, pp. 230–239, Nov. 2020, doi: 10.33633/joins.v5i2.3705.
- [3] I. N. Wardani, M. Ningsih, and R. Zusyana, "Pawitra Komunika jurnal komunikasi dan sosial humaniora PENGGUNAAN CLICKBAIT HEADLINE PADA PORTAL BERITA TRIBUNNEWS.COM," vol. 2, no. 1, 2021, [Online]. Available: <http://ejurnal.unim.ac.id/index.php/pawitrakomunika>
- [4] Y. Widhiyasana, T. Semiawan, I. Gibran, A. Mudzakir, and M. R. Noor, "Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification)," 2021.
- [5] J. Khatib Sulaiman, S. Rabbani, M. Khairul Anam, and I. Artikel Abstrak, "Optimalisasi Kinerja Klasifikasi Teks Berdasarkan Analisis Berbasis Aspek dan Model Hybrid Deep Learning," *Indonesian Journal of Computer Science*.
- [6] P. Studi Sistem Informasi, C. Guridno, A. Azimah, and S. Ningsih ABSTRAK, "Jurnal Sistem Informasi Bisnis (JUNSIBI) ANALISIS HYBRID METODE CNN DAN LSTM DALAM MEDIA BERITA ONLINE INDONESIA PENULIS 1)," vol. 5, no. 1, pp. 86–101, 1957, doi: 10.55122/junsibi.v5i1.1202.
- [7] M. R. Kertanegara, "Penggunaan Clickbait Headline pada Situs Berita," 2018.
- [8] N. Rahmatika, G. F. Prisanto, S. Tinggi Ilmu Komunikasi InterStudi, J. I. Wijaya No, and J. Selatan, "Pengaruh Berita Clickbait Terhadap Kepercayaan pada Media di Era Attention Economy."
- [9] J. Homepage, R. Rahman Salam, M. Fajri Jamil, and Y. Ibrahim, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," vol. 3, pp. 27–35, 2023.
- [10] R. Hayami, "KLASIFIKASI TEKS BERITA BERBAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING DAN DEEP LEARNING: STUDI LITERATUR," 2023. [Online]. Available: <https://ieeexplore.ieee.org/>
- [11] R. Nanda, E. Haerani, S. K. Gusti, and S. Ramadhani, "Klasifikasi Berita Menggunakan Metode Support Vector Machine," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, 2022.
- [12] D. Intan Af *et al.*, "Pengaruh Parameter Word2Vec terhadap Performa Deep Learning pada Klasifikasi Sentimen," vol. 6, no. 3, 2021.
- [13] I. Rifky Hendrawan, E. Utami, and A. D. Hartanto, "Analisis Perbandingan Metode Tf-Idf dan Word2vec pada Klasifikasi Teks Sentimen Masyarakat Terhadap Produk Lokal di Indonesia."
- [14] Adhitya Karel Maulaya and Junadhi, "Analisis Sentimen Menggunakan Support Vector Machine Masyarakat Indonesia Di Twitter Terkait Bjorka," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 495–500, Dec. 2022, doi: 10.37859/coscitech.v3i3.4358.
- [15] B. Arief, H. Kholifatullah, and A. Prihanto, "Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech," *Journal of Informatics and Computer Science*, vol. 04, 2023.
- [16] Y. Hendra, H. Mukhtar, and R. Hafsari, "PREDIKSI CURAH HUJAN DI KOTA PEKANBARU MENGGUNAKAN LSTM (LONG SHORT TERM MEMORY)," 2023.