

**ANALISIS HYBRID METODE CNN DAN LSTM
DALAM MEDIA BERITA ONLINE INDONESIA****PENULIS**¹⁾Ciptoningaji Guridno, ²⁾Ariana Azimah, ³⁾Sari Ningsih**ABSTRAK**

Berita palsu atau disinformasi merupakan ancaman serius dalam ekosistem media *online*. Penyebaran berita palsu dapat mengganggu informasi yang akurat dan dapat mempengaruhi masyarakat dan opini publik. Dalam penelitian ini, Penulis mengusulkan pendekatan hibrida yang mengintegrasikan Convolutional Neural Network (CNN) dan *Long Short-Term Memory* (LSTM) untuk menganalisis konten media berita *online* di Indonesia. Metode hibrid ini ditujukan untuk memahami dan menginterpretasikan dinamika informasi yang disampaikan melalui berita *online* dengan lebih efektif. Penulis mengumpulkan dan memproses *dataset* besar dari artikel berita *online* dalam Bahasa Indonesia, lalu menerapkan CNN untuk ekstraksi fitur teks dan LSTM untuk memodelkan sekuensialitas data dalam artikel. Hasil eksperimen menunjukkan bahwa model hibrid CNN-LSTM mampu meningkatkan akurasi klasifikasi topik berita dan sentiment analisis dibandingkan dengan metode standar. Penelitian ini memberikan wawasan baru tentang aplikasi pembelajaran mesin dalam media berita dan menawarkan metode yang inovatif untuk analisis teks pada skala besar.

Kata Kunci

CNN, LSTM, Media Berita Online, Analisis Teks, *Machine Learning***AFILIASI**

Program Studi

^{1,3)}Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika

Nama Institusi

²⁾Informatika, Fakultas Teknologi Komunikasi dan Informatika

Alamat Institusi

¹⁻³⁾Universitas Nasional¹⁻³⁾Jl. Sawo Manila No.61, RT.14/RW.7, Pejaten Bar., Ps. Minggu, Jakarta Selatan**KORESPONDENSI**

Penulis

Sari Ningsih

Email

lectures.sariningsih@gmail.com**LICENSE**

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

I. PENDAHULUAN

1.1 Latar Belakang

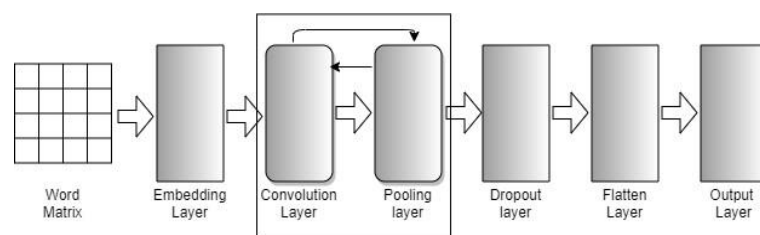
Di era digital yang kita alami saat ini, ketersediaan dan akses terhadap berita melalui media *online* menjadi semakin tak terbatas. Meskipun kemudahan akses ini memberikan keuntungan dalam penyebaran informasi, ia juga membuka peluang untuk penyebaran berita palsu atau *hoax* yang tersebar luas dengan kecepatan yang sama. Berita palsu tidak hanya menyesatkan publik tetapi juga memiliki potensi untuk memanipulasi opini publik dan keputusan politik, yang berdampak pada berbagai aspek sosial dan politik masyarakat.

Sampai sekarang, Indonesia telah meraih peringkat ketiga sebagai negara dengan jumlah pengguna internet terbanyak di Asia. Dilansir dari databoks.katadata.co.id, pada Januari 2023, jumlah pengguna internet di Indonesia mencapai 213 juta. Dampaknya terasa pada perilaku konsumsi masyarakat terutama dalam akses informasi digital, yang tercermin dari migrasinya banyak media cetak ke platform digital. Hasil survei UC Browser pada 2022 menunjukkan bahwa sekitar 95,4% pengguna internet di Indonesia membaca berita melalui *smartphone*, sementara media elektronik seperti TV hanya mencapai 45,9%, PC 15,3%, dan radio hanya sebesar 6,7%. [1]. Menurut Kementerian Komunikasi dan Informatika, sekitar 800.000 situs web di Indonesia tersebar informasi yang tidak benar. Kementerian tersebut melaporkan penyalahgunaan internet oleh sebagian individu dan kelompok dalam menyebarkan konten negatif yang mengakibatkan kecemasan dan ketidakpercayaan di tengah masyarakat. Antara 30 hingga 60 persen dari penduduk Indonesia terpapar informasi yang tidak valid ketika berinteraksi di ranah digital, sementara hanya 21-36 persen yang menyadari akan keberadaan penipuan semacam itu.[2][3]. Pendekatan Hybrid CNN-LSTM khususnya dianggap berpotensi karena menggabungkan kekuatan CNN dalam mengidentifikasi fitur-fitur lokal dalam data teks dengan kemampuan LSTM untuk memahami konteks dalam sekuens yang lebih panjang, yang esensial dalam memahami nuansa dalam narasi [4][10]. Pendekatan ini dipilih karena telah terbukti efektif dalam tugas-tugas terkait seperti analisis sentimen dan klasifikasi teks, yang memiliki kemiripan dengan masalah deteksi berita palsu[5]. Melalui penggunaan model Hybrid CNN-LSTM, penelitian ini bertujuan untuk menciptakan sebuah sistem yang dapat tidak hanya membedakan fakta dari fiksi dengan akurasi yang lebih tinggi tetapi juga memproses volume data yang besar secara *real-time*, yang merupakan tantangan utama dalam pendeteksian berita palsu.

1.2 Tinjauan Pustaka

1.2.1 CNN

CNN adalah algoritma *deep learning* yang sangat efektif untuk analisis gambar. *CNN* menggunakan lapisan konvolusi yang otomatis dan adaptif belajar untuk mendeteksi fitur penting dari gambar untuk tugas seperti klasifikasi atau deteksi objek.



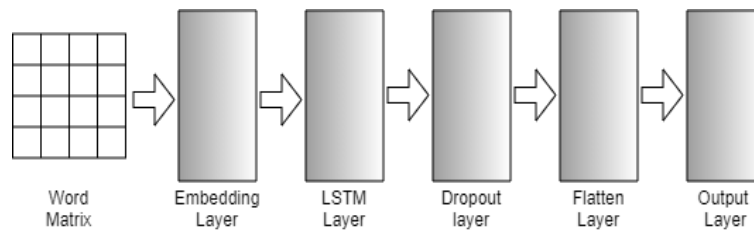
Gambar 1. Arsitektur Model CNN

Pada gambar1, Arsitektur Model CNN terdiri dari tiga jenis lapisan: lapisan masukan, lapisan tersembunyi, dan lapisan keluaran. Lapisan tersembunyi dalam jaringan saraf terdiri dari lapisan konvolusi, di mana kernel konvolusi bergerak melalui matriks masukan untuk membentuk peta fitur yang menjadi masukan untuk lapisan berikutnya. Kemudian, ditambahkan lapisan-lapisan lain seperti lapisan pooling, lapisan yang sepenuhnya terhubung, dan lapisan normalisasi[6].

1.2.2 LSTM

Long Short-Term Memory (LSTM) adalah variasi dari *Recurrent Neural Network (RNN)*. Jika RNN hanya bisa menyimpan informasi terkini, *LSTM* dapat menangani data jangka panjang. Selain itu, model *RNN*

mengalami masalah yang dikenal sebagai masalah gradien yang menghilang saat memproses data urutan Panjang, namun, *LSTM* dapat menghindari tantangan ini selama pelatihan. Model ini dapat mengingat data seri waktu jangka panjang sebelumnya dan memungkinkan kontrol otomatis dalam keadaan sel untuk mempertahankan sifat yang berguna dan membuang yang tidak relevan[7].



Gambar 2. Arsitektur Model LSTM

Pada Gambar 2, ada tiga gerbang yang mengatur fitur dalam model *LSTM*: gerbang masuk, gerbang lupa, dan gerbang keluar. Gerbang masuk berfungsi untuk membiarkan data baru masuk ke dalam keadaan sel[8]. Gerbang lupa menghapus informasi yang tidak relevan dari keadaan sel. Gerbang keluar, yang mengontrol informasi yang diambil dari keadaan sel, kemudian menentukan apa yang akan menjadi keadaan tersembunyi selanjutnya. Model *LSTM* dapat secara otomatis menyimpan atau menghapus memori yang tercatat menggunakan gerbang-gerbang ini.

II. METODE PENELITIAN

2.1 Penentuan Subjek Penelitian

Subjek penelitian yang terlibat dalam penelitian ini adalah online news media di Indonesia. Dengan mengambil dari 3 media besar CNN, Kompas dan tempo sebagai acuan data yang akan diolah dalam penelitian kali ini untuk mendapatkan jumlah berita hoax politik di Indonesia. Data bukti data hoax didapat dari Turnbackhoax yang digunakan sebagai data pengujian berita hoax untuk mendapatkan hasil akurasi dari penggunaan Metode Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM)[9].

2.2 Fokus Penelitian

Fokus Dalam penelitian ini, fokus utamanya adalah menerapkan metode *hybrid Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)* dalam menganalisis media berita *online* di Indonesia. Penulis mengumpulkan dan memproses dataset besar dari artikel berita *online* dalam bahasa Indonesia. *CNN* digunakan untuk ekstraksi fitur teks, sedangkan *LSTM* digunakan untuk memodelkan sekuensialitas data dalam artikel. Hasil eksperimen menunjukkan bahwa model *hybrid CNN-LSTM* mampu meningkatkan akurasi klasifikasi topik berita dan analisis sentimen dibandingkan dengan metode standar.

2.3 Sumber Data

Data yang digunakan dalam penelitian ini berasal dari *online news* media di Indonesia. Dataset yang digunakan mencakup berbagai artikel berita dalam bahasa Indonesia yang diambil dari website Kaggle dan penambahan data dari website CNN, Kompas dan Tempo melalui data *scraping* dengan total jumlah data 20.000 artikel. Dalam penelitian ini, data yang digunakan terbagi menjadi dua jenis data, yaitu data uji dan data pelatihan, dimana data uji berisikan artikel dari media informasi CNN, Kompas, dan Tempo, Sedangkan data pelatihan dari *TurnbackHoax*

Tabel 1. Jumlah Artikel Sumber Data

Jumlah Artikel			
CNN	KOMPAS	Tempo	TurnbackHoax
10.000 Artikel	4.745 Artikel	6.592 Artikel	10.253 Artikel
Total 31590 Artikel			

2.4 Teknik Pengumpulan Data

Dalam penelitian ini, data dikumpulkan melalui dua metode utama untuk memastikan keragaman dan kualitas data. Pertama, *dataset* yang berkaitan dengan teks berita telah diunduh dari platform *Kaggle*, yang dikenal sebagai repositori data yang luas dan sering digunakan oleh komunitas ilmu data. Langkah-langkah web *scraping* pada sebuah *website* adalah sebagai berikut terlihat pada Gambar 3.

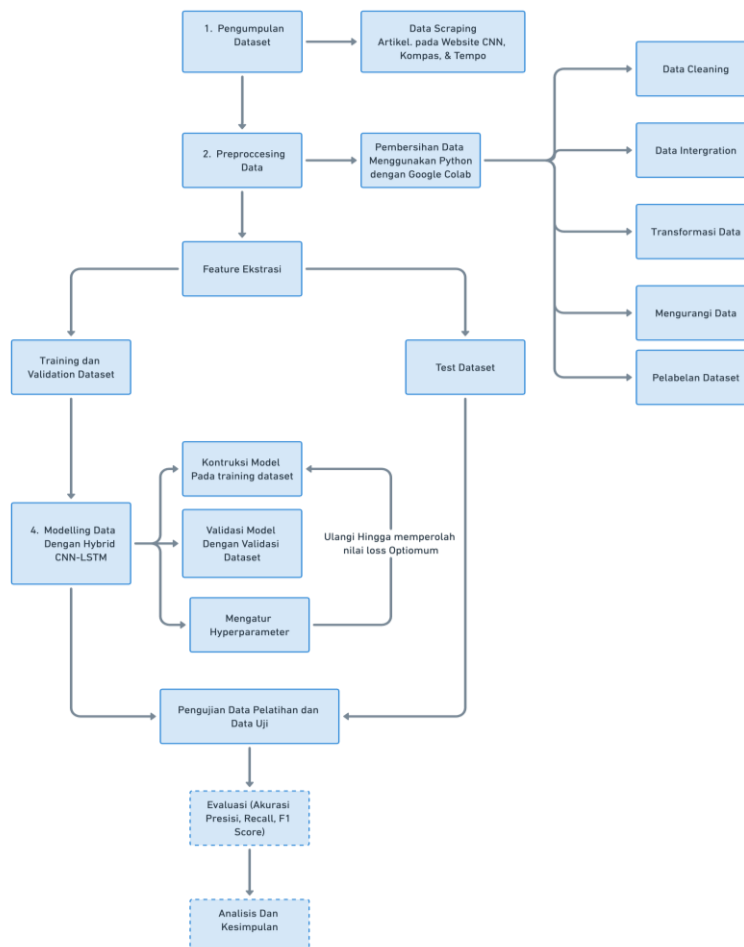


Gambar 3. Tahapan Web Scraping Data

- 1) Penentuan Target *Scraping*: Tentukan data spesifik yang ingin di ekstrak dari *website*, pada *case* ini fokus utama.
- 2) Inspeksi *Website*: Menggunakan *tools* pengembangan browser untuk memeriksa struktur HTML dari *website*. Identifikasi elemen-elemen HTML (seperti *div*, *table*, *span*) yang berisi data target.
- 3) Pemrograman *Script Scraping*: membuat *script* menggunakan bahasa pemrograman seperti *Python*. Dalam penelitian ini menggunakan pustaka seperti *requests* untuk mengirim permintaan HTTP ke *website* dan *BeautifulSoup* untuk mengurai HTML dan mengambil data.
- 4) Eksekusi *Script* dan Ekstraksi Data: Jalankan *script* untuk mengirim permintaan ke *website*, menerima respons *HTML*, dan mengurai *HTML* tersebut untuk mengekstrak data yang diinginkan.
- 5) Penyimpanan Data: Simpan data yang telah diekstrak ke dalam format yang sesuai, seperti CSV.

2.5 Desain Pengolahan Model Hybrid CNN-LSTM

Pada Gambar 4. Desain Pengolahan Model ini menggunakan metode penelitian hybrid CNN-LSTM. Langkah-langkah penelitian melibatkan pengumpulan data, preprocessing data, ekstraksi fitur, dan pengolahan data menggunakan model hybrid CNN-LSTM.



Gambar 4. Desain Pengolahan Model

2.6 Analisis dan Kesimpulan Pengolahan Data

Sistem pengolahan data dalam penelitian ini memiliki tujuan utama untuk melakukan klasifikasi topik berita dan analisis sentimen. Metode yang digunakan adalah *hybrid CNN-LSTM* yang melibatkan ekstraksi fitur teks dan pemodelan data sekuensial. Evaluasi kinerja sistem dilakukan dengan menggunakan metrik seperti *akurasi*, *presisi*, *recall*, dan *F1 score*. Tantangan yang mungkin dihadapi adalah keterbatasan sumber daya komputasi dan keandalan data. Untuk meningkatkan sistem, rekomendasi yang dapat dilakukan adalah menggunakan teknik pengumpulan data yang lebih efisien dan melakukan pengoptimalan model untuk meningkatkan akurasi.

III. HASIL DAN PEMBAHASAN

3.1 Web Scrapping Data

Penelitian ini mengumpulkan total 31.590 artikel dari empat sumber berita terpercaya di Indonesia: CNN Indonesia, Kompas, Tempo, dan TurnBackHoax. Proses pengumpulan data dilakukan melalui teknik *web scrapping*, yang memungkinkan pengambilan konten secara otomatis dari situs web. Artikel dipilih berdasarkan kata kunci yang relevan dengan topik penelitian, mencakup beragam isu seperti politik, kesehatan, teknologi, dan lingkungan, terlihat pada Gambar 5, 6 dan 7 Data *Output Hasil Proses Web Scrapping CNN & Kompas*.

```

from Scrapper import run_cnn, run_kompas, run_tempo, run_turnbackhoax

# run_cnn(num_of_page=1)
# run_kompas(num_of_page=449)
# run_tempo(num_of_day=400, start_date='2022-01-01')
run_turnbackhoax(num_of_page=520)

*** Starting Turnbackhoax.id News Scrapper
Processing News Pages:: 58% | 300/520 [15:34-12:35, 3.43s/it]
    
```

Gambar 5. Python Web Scrapping

index	title	raw_timestamp	original	tags	author	url	cleaned_label	timestamp	cleaned_token_length	summarized
0	[BALAH] Anies Baswedan Dekat Dengan Aliran Kru...	Maret 1, 2023	Hasil Periksa Fakta Gabriela Nauli Sinaga (Un...)	FinaH/HasutHoax	Pemeriksa Fakta Junior	https://turnbackhoax.id/2023/03/01/salah-anies...	1	2023-03-01	43	ONTA YAMAN NGGAK PEDULI ITU APA YANG PENTING D...
1	[BALAH] Hakim Wahyu Iman Santoso Alami Kecelakaan...	Maret 1, 2023	Hasil Periksa Fakta Gabriela Nauli Sinaga (Un...)	FinaH/HasutHoax	Pemeriksa Fakta Junior	https://turnbackhoax.id/2023/03/01/salah-hakim...	0	2023-03-01	36	Ya Allah gimna keadaan pa hakin? Ini befer gas...

index	title	raw_timestamp	original	tags	author	url	cleaned_label	timestamp	cleaned_token_length	summarized
0	Efek Ekor Jaa Pencalonan Anies, Elektabilitas...	21 Februari 2023, 15:35 WIB	Hasil jajak pendapat yang diselenggarakan Lib...	Survei Litbang Kompas/Elektabilitas Nasdem Nal...	NaN	https://video.kompas.com/watch/258152/efek-eko...	0	2023-02-21 15:30:00	118	Hasil jajak pendapat yang diselenggarakan Lib...
1	Ekonomi 2024 Ditargetkan Tumbuh 5,7 Persen, pa...	Kompas.com - 21/02/2023, 14:22 WIB	JAKARTA, KOMPAS.com - Pemerintah menargetkan p...	Jakarta,Ekonomi 2024	Penulis: Yohana Artha Uyi Editor: Aprilia Ika	http://money.kompas.com/read/2023/02/21/142238...	0	2023-02-21 14:22:00	423	Pemerintah menargetkan pertumbuhan ekonomi 6...

Gambar 6. Data Output Hasil Proses Web Scrapping CNN & Kompas

index	title	raw_timestamp	original	tags	author	url	cleaned_label	timestamp	cleaned_token_length	summarized
0	Ma'Yuf Amin akan Sakaikan Lagi Timnas Indonesia...	Sabtu, 1 Januari 2022 17:16 WIB	TEMPO.CO, Jakarta - Wakil Presiden Ma'ruf Amin...	Ma'Yuf Amin,Pala AFF 2023,Indonesia vs Thaila...	Reporter Egi Adiyatama Editor Adhya Budiman	https://nasional.tempo.co/read/1545904/ma-yuf...	0	2022-01-01 17:14:00	41	Timnas Indonesia Vs Thailand akan Tangpi Lagi...
1	Menag Yaqut Canangkan 2022 Sebagai Tahun Tol...	Sabtu, 1 Januari 2022 15:05 WIB	TEMPO.CO, Jakarta - Menteri Agama Yaqut Cholli...	Menag,Yaqut Cholli Qomari,Gas Yaqut,Toleransi...	Reporter Egi Adiyatama Editor Adhya Budiman	https://nasional.tempo.co/read/1545477/menag-y...	0	2022-01-01 15:05:00	209	Pemerintah menacangkan 2022 sebagai Tahun Tol...

index	title	raw_timestamp	original	tags	author	url	cleaned_label	timestamp	cleaned_token_length	summarized
0	[BALAH] Anies Baswedan Dekat Dengan Aliran Kru...	Maret 1, 2023	Hasil Periksa Fakta Gabriela Nauli Sinaga (Un...)	FinaH/HasutHoax	Pemeriksa Fakta Junior	https://turnbackhoax.id/2023/03/01/salah-anies...	1	2023-03-01	43	ONTA YAMAN NGGAK PEDULI ITU APA YANG PENTING D...
1	[BALAH] Hakim Wahyu Iman Santoso Alami Kecelakaan...	Maret 1, 2023	Hasil Periksa Fakta Gabriela Nauli Sinaga (Un...)	FinaH/HasutHoax	Pemeriksa Fakta Junior	https://turnbackhoax.id/2023/03/01/salah-hakim...	0	2023-03-01	36	Ya Allah gimna keadaan pa hakin? Ini befer gas...

Gambar 7. Data Output Hasil Proses Web Scrapping Tempo & Turnbackhoax

3.2 Pra-pemrosesan Data

Langkah awal pra-pemrosesan terlihat pada Gambar 8 s.d 15, melibatkan penghapusan elemen non-teksual seperti gambar dan tautan, diikuti oleh penghapusan karakter khusus dan normalisasi teks (misalnya, mengubah semua huruf menjadi huruf kecil). Tokenisasi dilakukan untuk mengubah teks menjadi sekumpulan token yang lebih kecil, memudahkan model untuk memproses informasi. Data kemudian dibagi menjadi set pelatihan (70%), validasi (15%), dan pengujian (15%) untuk evaluasi model.

```
Drop NaN data value

[9] df_cnn = df_cnn[df_cnn['cleaned'].isna() == False]
df_kompas = df_kompas[df_kompas['cleaned'].isna() == False]
df_tempo = df_tempo[df_tempo['cleaned'].isna() == False]
df_turnbackhoax = df_turnbackhoax[df_turnbackhoax['cleaned'].isna() == False]
```

Gambar 8. Membuang Nilai Kosong Pada 4 Data CSV

```
Convert semua nilai data menjadi string

[10] df_cnn['cleaned'] = df_cnn['cleaned'].astype(str)
df_kompas['cleaned'] = df_kompas['cleaned'].astype(str)
df_tempo['cleaned'] = df_tempo['cleaned'].astype(str)
df_turnbackhoax['cleaned'] = df_turnbackhoax['cleaned'].astype(str)

<ipython-input-10-929bb42dda9a>:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_kompas['cleaned'] = df_kompas['cleaned'].astype(str)
<ipython-input-10-929bb42dda9a>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
df_turnbackhoax['cleaned'] = df_turnbackhoax['cleaned'].astype(str)
```

Gambar 9. Mengubah Tipe Data Menjadi String

```
Pemberian Tanda Baca pada setiap kolom news

[11] def space_to_punct(text):
    for punct in string.punctuation:
        text = text.replace(punct, f' {punct} ')

    text = re.sub(' +', ' ', text)
    return text

df_cnn['cleaned'] = df_cnn['cleaned'].map(lambda x: space_to_punct(x))
df_kompas['cleaned'] = df_kompas['cleaned'].map(lambda x: space_to_punct(x))
df_tempo['cleaned'] = df_tempo['cleaned'].map(lambda x: space_to_punct(x))
df_turnbackhoax['cleaned'] = df_turnbackhoax['cleaned'].map(lambda x: space_to_punct(x))
```

Gambar 10. Menambahkan Tanda Baca Pada Kolom News

```
Mengubah Dataset baru dengan mengambil text yang telah melalui preprocessing data

df_cnn_cleaned = df_cnn[['cleaned']]
df_cnn_cleaned['labels'] = 0
df_kompas_cleaned = df_kompas[['cleaned']]
df_kompas_cleaned['labels'] = 0
df_tempo_cleaned = df_tempo[['cleaned']]
df_tempo_cleaned['labels'] = 0
df_turnbackhoax_cleaned = df_turnbackhoax[['cleaned']]
df_turnbackhoax_cleaned['labels'] = 1
```

Gambar 11. Eskport Data Baru yang Sudah Melalui Preprocessing Data

```
df_cnn_cleaned.head(2)
```

index	cleaned	labels
0	Anies di Milad BKMT : Pengajian Menghasilkan I...	0
1	Edy Soal Pilgub Sumut : Kalau yang Maju Abal -...	0

```
[14] df_kompas_cleaned.head(2)
```

index	cleaned	labels
0	Efek Ekor Jas Pencalonan Anies , Elektabilitas...	0
1	Ekonomi 2024 Ditargetkan Tumbuh 5 , 7 Persen , ...	0

```
[15] df_tempo_cleaned.head(2)
```

index	cleaned	labels
0	Ma ' ruf Amin akan Saksikan Lagi Timnas Indone...	0
1	Menag Yaqut Canangkan 2022 Sebagai Tahun Toler...	0

```
[16] df_turnbackhoax_cleaned.head(2)
```

index	cleaned	labels
0	BISA DILIHAT SI ONTA YAMAN NGGAK PEDULI ITU AP...	1
1	ini bener gasih ? ? Ya Allah gimna keadaan pa ...	1

Gambar 12. Pengecekan Nilai Data Pada Setiap Data

```
[17] dataset_cnn_train_test = Dataset.from_pandas(df_cnn_cleaned).train_test_split(seed=1, test_size=0.2).remove_columns(["index"])
dataset_kompas_train_test = Dataset.from_pandas(df_kompas_cleaned).train_test_split(seed=1, test_size=0.2).remove_columns(["index"])
dataset_tempo_train_test = Dataset.from_pandas(df_tempo_cleaned).train_test_split(seed=1, test_size=0.2).remove_columns(["index"])
dataset_turnbackhoax_train_test = Dataset.from_pandas(df_turnbackhoax_cleaned).train_test_split(seed=1, test_size=0.2).remove_columns(["index"])

[18] dataset_turnbackhoax_train_test
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 5201
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 1301
  })
})

dataset_cnn_train_test
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 7704
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 1926
  })
})

[20] dataset_kompas_train_test
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 3778
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 945
  })
})

[21] dataset_tempo_train_test
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 5273
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 1319
  })
})
```

Gambar 13. Pembagian Data Uji & Latihan

```
[22] dataset_cnn_test_eval = dataset_cnn_train_test['test'].train_test_split(seed=1, test_size=0.5)
dataset_kompas_test_eval = dataset_kompas_train_test['test'].train_test_split(seed=1, test_size=0.5)
dataset_tempo_test_eval = dataset_tempo_train_test['test'].train_test_split(seed=1, test_size=0.5)
dataset_turnbackhoax_test_eval = dataset_turnbackhoax_train_test['test'].train_test_split(seed=1, test_size=0.5)

[23] dataset_turnbackhoax_test_eval
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 650
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 651
  })
})

dataset_cnn_test_eval
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 963
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 963
  })
})

[25] dataset_kompas_test_eval
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 472
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 473
  })
})

[26] dataset_tempo_test_eval
DatasetDict({
  train: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 659
  })
  test: Dataset({
    features: ['cleaned', 'labels'],
    num_rows: 660
  })
})
```

Gambar 14. Pembagian Dataset Eval

```

Penggabungan Data Test, Evaluasi dan Pelatihan

ds_train = concatenate_datasets([
    dataset_cnn_train_test['train'],
    dataset_kompas_train_test['train'],
    dataset_tempo_train_test['train'],
    dataset_turnbackhoax_train_test['train']
])
ds_eval = concatenate_datasets([
    dataset_cnn_test_eval['train'],
    dataset_kompas_test_eval['train'],
    dataset_tempo_test_eval['train'],
    dataset_turnbackhoax_test_eval['train']
])
ds_test = concatenate_datasets([
    dataset_cnn_test_eval['test'],
    dataset_kompas_test_eval['test'],
    dataset_tempo_test_eval['test'],
    dataset_turnbackhoax_test_eval['test']
])

[28] ds_train
Dataset({
  features: ['cleaned', 'labels'],
  num_rows: 21956
})

[29] ds_eval
Dataset({
  features: ['cleaned', 'labels'],
  num_rows: 2744
})

[30] ds_test
Dataset({
  features: ['cleaned', 'labels'],
  num_rows: 2747
})

```

Gambar 15. Penggabungan Data Train, Test & Eval Menjadi Dataset

3.3 Transformer IndoBERT

Model yang digunakan adalah BERT (Bidirectional Encoder Representations from Transformers), khususnya varian IndoBERT untuk adaptasi bahasa Indonesia. IndoBERT dipilih karena pemahamannya yang baik terhadap nuansa bahasa Indonesia dan telah terbukti efektif dalam berbagai tugas pemrosesan bahasa alami. Model ini dilatih dengan parameter berikut: ukuran batch 32, laju pembelajaran $2e-5$, dan dilakukan selama 5 epoch. Optimasi dilakukan menggunakan AdamW dengan weight decay untuk menghindari *overfitting*, terlihat pada gambar 16 dan 17.

```

Penggabungan Proses data menggunakan Cuda GPU dari CPU saja

(31) device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

Tokenizer Menggunakan IndoBert

(32) tokenizer = AutoTokenizer.from_pretrained("indobenchmark/indobert-base-p1")
model = AutoModelForSequenceClassification.from_pretrained("indobenchmark/indobert-base-p1", num_labels=5)

/usr/local/lib/python3.10/dist-packages/huggingface_hub/utils/_token.py:88: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab (https://huggingface.co/settings/tokens), set it as secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access public models or datasets.
warnings.warn

tokenizer.config.json: 100% |#####| 2.00/2.00 [00:00<00:00, 1808B/s]
config.json: 100% |#####| 1.53K/1.53K [00:00<00:00, 141KB/s]
vocab.txt: 100% |#####| 229K/229K [00:00<00:00, 1.69MB/s]
special_tokens_map.json: 100% |#####| 112/112 [00:00<00:00, 9.51KB/s]
pytorch_model.bin: 100% |#####| 498M/498M [00:01<00:00, 266MB/s]

Some weights of BertForSequenceClassification were not initialized from the model checkpoint at indobenchmark/indobert-base-p1 and are newly initialized: ['classifier.bias', 'classifier.weight']
You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

Proses Tokenize

def tokenize_dataset(data):
    # Keys of the returned dictionary will be added to the dataset as columns
    return tokenizer(data["cleaned"], padding=True, truncation=True, max_length=512)

ds_train = ds_train.map(tokenize_dataset)
ds_eval = ds_eval.map(tokenize_dataset)
ds_test = ds_test.map(tokenize_dataset)

Map: 100% |#####| 21956/21956 [00:28<00:00, 1036.56 examples/s]
Map: 100% |#####| 2744/2744 [00:03<00:00, 947.34 examples/s]
Map: 100% |#####| 2747/2747 [00:03<00:00, 1106.57 examples/s]

```

Gambar 16. Proses Tokenizer Menggunakan IndoBERT

```

(34) ds_train_tokenized = ds_train.remove_columns(['cleaned'])
ds_eval_tokenized = ds_eval.remove_columns(['cleaned'])
ds_test_tokenized = ds_test.remove_columns(['cleaned'])

dataset_full = DatasetDict({
    'train': ds_train_tokenized,
    'test': ds_eval_tokenized,
    'eval': ds_test_tokenized
})

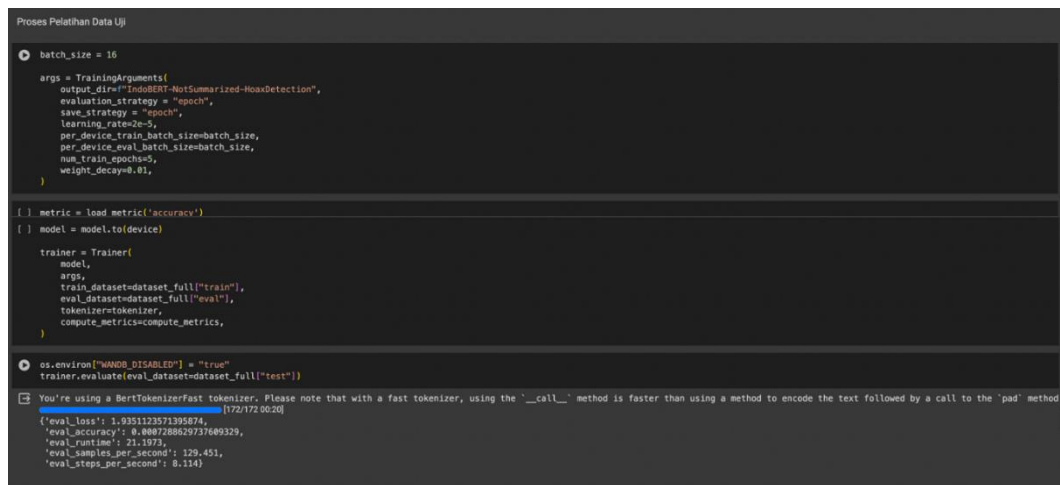
dataset_full
DatasetDict({
  train: Dataset({
    features: ['labels', 'input_ids', 'token_type_ids', 'attention_mask'],
    num_rows: 21956
  })
  test: Dataset({
    features: ['labels', 'input_ids', 'token_type_ids', 'attention_mask'],
    num_rows: 2744
  })
  eval: Dataset({
    features: ['labels', 'input_ids', 'token_type_ids', 'attention_mask'],
    num_rows: 2747
  })
})

```

Gambar 17. Proses Tokenize Dataset

3.4 Arsitektur mode Pelatihan Data Uji

Arsitektur model meliputi beberapa lapisan, seperti *embedding layer*, LSTM layer, *dropout layer*, dan *dense layer*, dirancang untuk mencapai performa optimal. Ukuran *embedding* yang digunakan adalah 100 dengan panjang *input* sebanyak 1.135, dan LSTM layer menggunakan 16 neuron. Pengaturan ini dipilih berdasarkan eksperimen dan evaluasi untuk menghasilkan model yang mampu memprediksi dengan akurat.



```

Proses Pelatihan Data Uji

batch_size = 16

args = TrainingArguments(
  output_dir='IndoBERT-NotSummarized-HoaxDetection',
  evaluation_strategy = "epoch",
  save_strategy = "epoch",
  learning_rate=2e-5,
  per_device_train_batch_size=batch_size,
  per_device_eval_batch_size=batch_size,
  num_train_epochs=5,
  weight_decay=0.01,
)

metric = load_metric('accuracy')
model = model.to(device)

trainer = Trainer(
  model,
  args,
  train_dataset=dataset_full["train"],
  eval_dataset=dataset_full["eval"],
  tokenizer=tokenizer,
  compute_metrics=compute_metrics,
)


os.environ["WANDB_DISABLED"] = "true"
trainer.evaluate(eval_dataset=dataset_full["test"])

You're using a BertTokenizerFast tokenizer. Please note that with a fast tokenizer, using the '_call_' method is faster than using a method to encode the text followed by a call to the 'pad' method

{'eval_loss': 1.9351123571395874,
 'eval_accuracy': 0.9978045481849353,
 'eval_runtime': 21.1973,
 'eval_samples_per_second': 129.451,
 'eval_steps_per_second': 0.134}

```

Gambar 18. Parameter Pelatihan Data Uji



```

Pelatihan Data Uji sebagai dasar pengetahuan Algoritma

# train model
trainer.train()

[6865/6865 43.09, Epoch 5/5]

Epoch Training Loss Validation Loss Accuracy
1 0.033900 0.021213 0.993811
2 0.008000 0.032729 0.993083
3 0.003900 0.016167 0.997816
4 0.001700 0.016247 0.997452
5 0.000200 0.016214 0.997452

TrainOutput(global_step=6865, training_loss=0.013785433763731746, metrics={'train_runtime': 2598.7875, 'train_samples_per_second': 42.375, 'train_steps_per_second': 2.65, 'total_flos': 2.845111685988843e+16, 'train_loss': 0.013785433763731746, 'epoch': 5.0})

```

Gambar 19. Proses Pelatihan Data Uji



```

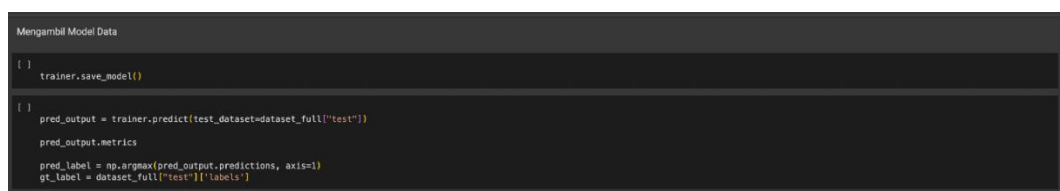
Evaluasi Data

trainer.evaluate(eval_dataset=dataset_full["test"])

{'eval_loss': 0.81972723193466634,
 'eval_accuracy': 0.9978045481849353,
 'eval_runtime': 20.3483,
 'eval_samples_per_second': 134.985,
 'eval_steps_per_second': 0.456,
 'epoch': 5.0}

```

Gambar 20. Pengujian Data Train dengan Data Eval



```

Mengambil Model Data

trainer.save_model()

pred_output = trainer.predict(test_dataset=dataset_full["test"])
pred_output.metrics

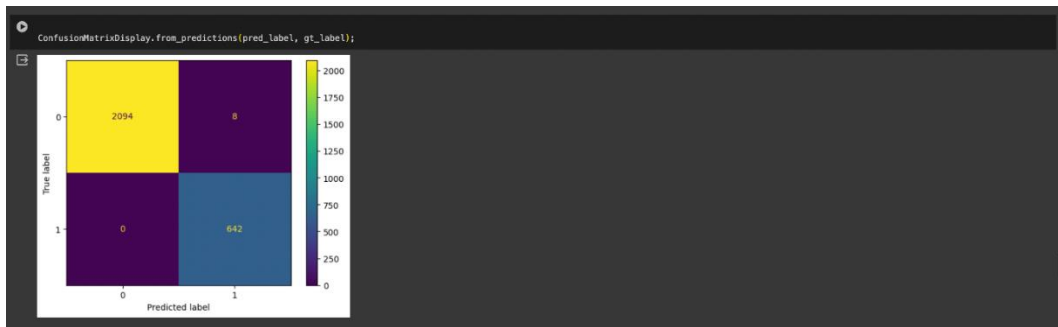
pred_label = np.argmax(pred_output.predictions, axis=-1)
gt_label = dataset_full["test"]['labels']

```

Gambar 21. Save model Parameter

3.5 Hasil Pengujian Data Latihan Dengan Data Uji

Pada Gambar 22 s.d 24, Evaluasi model dilakukan menggunakan data uji yang terdiri dari 1.097 artikel, dengan 805 di antaranya berlabel Hoax dan 292 berlabel Non-Hoax. Tujuan utama dari evaluasi ini adalah untuk menguji sejauh mana model dapat diandalkan dalam memprediksi kelas dari data yang diberikan. Hasil evaluasi menunjukkan bahwa model berhasil memprediksi 837 artikel Hoax dengan benar dan hanya 3 kesalahan prediksi, sedangkan untuk artikel Non-Hoax, model berhasil memprediksi semua 257 artikel dengan benar, tanpa kesalahan prediksi.



Gambar 22 .Confusion Matrix Hasil Data Uji Dengan Data Latihan

```

# take it from tokenized data instead of input
test_data = tokenizer.batch_decode(dataset_full["test"]["input_ids"], skip_special_tokens=True)

# create gt and prediction comparison
pred_compare = []
for i, news in enumerate(test_data):
    pred_compare.append((news, gt_label[i], pred_label[i]))

# turn it to dataframe for easy processing
df_pred_compare = pd.DataFrame(pred_compare)
df_pred_compare = df_pred_compare.rename(columns={0:'news', 1:'labels', 2:'prediction'})
df_pred_compare.to_csv('predictions2.csv')
df_pred_compare

# Setelah membuat DataFrame
df_pred_compare = pd.DataFrame(pred_compare, columns=['news', 'labels', 'prediction'])

# Menyetel ulang indeks dan mengganti nama kolom indeks menjadi 'no'
df_pred_compare.reset_index(inplace=True)
df_pred_compare = df_pred_compare.rename(columns={'index': 'no'})

# Menyimpan ke CSV (optional)
df_pred_compare.to_csv('predictions.csv', index=False) # index=False untuk tidak menyertakan kolom indeks dalam file CSV

# Menampilkan DataFrame
df_pred_compare

```

no	news	labels	prediction
0	0 waktu fit and proper test andika jadi panglima...	0	0
1	1 demokrat : masa depan buruh dikubur uu cipta k...	0	0
2	2 jokowi lantik kepala otorita iin pekan depan, ...	0	0
3	3 komisi x minta ke jokowi: jangan ada rektor...	0	0
4	4 pks sambut baik anies jadi capres nasdem : sim...	0	0
...
2739	2739 farel prayoga meninggal dunia sakit? kabar ter...	1	1

Gambar 23. Perbandingan Data Train Dengan Data Test

```

df_wrong_pred = df_pred_compare[df_pred_compare['labels'] != df_pred_compare['prediction']]
df_pred_compare.to_csv('predictions-wrong.csv')

def label_to_str(label_int):
    label = 'Hoax' if label_int == 1 else 'Fakta'
    return f'({label}) {label_int}'

for i, row in enumerate(df_wrong_pred.values):
    print(f'{i+1} | {row[0]}')
    print(f'=> Ground Truth label: {label_to_str(row[1])}, but predicted as: {label_to_str(row[2])}')
    print()

```

1 2155	>> Ground Truth label: Fakta [agenda pbb new world order united nations agenda 21 / 2030 mission goalis. satu pemerintahan dunia, dunia tanpa uang tunai, satu bank sentral dunia, satu militer dunia.
2 2232	>> Ground Truth label: Fakta [jokowi sim keliling polres blitar januari 2022 1. senin, 3 januari 2022 di kantor desa sumberjati kademangan 2. selasa, 4 januari 2022 di kantor desa ngembul binangun 3.
3 2267	>> Ground Truth label: Fakta [megawati minta sri mulyani jangan pelit soal dana wakaf untuk melunasi utang ke tiongkok], but predicted as: Hoax [1]
4 2270	>> Ground Truth label: Fakta [survei populi center : 71, 8 persen masyarakat bangga diplinin jokowi. hanya kardun dungu yang tak bangga akan presiden yang amanah ini], but predicted as: Hoax [1]
5 2334	>> Ground Truth label: Fakta [jokowi lepas baju pdip, megawati marah jokowi tak mau lakukan ini], but predicted as: Hoax [1]
6 2425	>> Ground Truth label: Fakta [ini penyusup rombongan ke acara jokowi, dia bawa bon di dalam kereta, sudah diturunkan di semarang dan dibawa ke polda untuk investigasi lebih lanjut.], but predicted as:
7 2438	>> Ground Truth label: Fakta [gempa m 6, 7 guncang banten, terasa hingga jakarta, bogor, depok], but predicted as: Hoax [1]
8 2487	>> Ground Truth label: Fakta [presiden as menjemput pm singapura, sementara jokowi hanya dijemput dubes indonesia untuk as. ada apa?], but predicted as: Hoax [1]

Gambar 24. Perbandingan Data Train Dengan Data Test 2

3.6 Klasifikasi Berita Hoax

Setelah mendapatkan parameter terbaik yaitu jumlah neuron sebanyak 16 dan jumlah batch size sebanyak 16 selanjutnya model akan di evaluasi. Pada tahap evaluasi, model akan diuji untuk memprediksi kelas data dengan menggunakan data uji yang berjumlah 4.391 dengan jumlah data Hoax sebanyak 1.040 dan jumlah data Non-Hoax 3.351. Tujuan evaluasi uji data terlihat pada table 1. untuk melihat seberapa besar model dapat dipercaya dalam memprediksi kelas.

Tabel 1 Hasil Prediksi dan Data Aktual

Aktual		Prediksi		Prediksi Meleset	
Hoax	Non Hoax	Hoax	Non Hoax	Hoax	Non Hoax
1.040	3.351	1.033	3.358	7	7

Hasil pengujian model ditunjukkan pada tabel 1 dimana berita yang berlabel *hoax* pada data aktual sebanyak 1.040 berita dan berita *non hoax* pada data aktual sebanyak 3.351 berita. Sedangkan pada data prediksi untuk berita berlabel *hoax* diprediksi sebanyak 1.033 berita dan berita *non hoax* 3.358 berita. Hal ini dapat dilihat terdapat 7 *data hoax* dan 7 *data non hoax* yang salah pada prediksinya yang menghasilkan tingkat akurasi 0.999932 dengan *validation loss* sebesar 0,00067. Hasil klasifikasi berita *hoax* yang telah dilakukan tercantum dalam Tabel 2 sebagai berikut:

Tabel 2. Hasil Klasifikasi Berita

Aktual	Prediksi	Berita Sebelum <i>Preprocessing</i>	Berita Setelah <i>Preprocessing</i>
Hoax	Hoax (TP(1,1))	<p>“Rusia menjadi negara pertama di dunia yang melakukan otopsi (post mortem) terhadap jenazah Covid-19. Setelah dilakukan penyelidikan menyeluruh, ditemukan bahwa Covid-19 tidak ada dalam bentuk virus, melainkan bakteri yang telah terpapar radiasi dan menggumpal melalui darah hingga menyebabkan kematian.</p> <p>Penyakit Covid-19 telah ditemukan menyebabkan pembekuan darah, yang menyebabkan pembekuan darah manusia dan pembekuan darah vena, yang membuat orang sulit bernapas karena otak, jantung, dan paru-paru tidak dapat menyerap oksigen, menyebabkan orang mati dengan cepat.</p> <p>Guna mengetahui penyebab kurangnya energi pernapasan, dokter Rusia tidak mendengarkan kesepakatan WHO, melainkan melakukan otopsi terhadap COVID-19. Setelah dokter membuka lengan, kaki, dan bagian tubuh lainnya dan memeriksanya dengan cermat, mereka menemukan bahwa pembuluh darah melebar dan berisi gumpalan darah, yang menghalangi aliran darah dan mengurangi aliran oksigen. Hal tersebut dapat menyebabkan kematian pada tubuh. Setelah mengetahui penelitian tersebut, Kementerian Kesehatan Rusia segera mengubah rencana pengobatan Covid-19 dan menggunakan aspirin untuk pasien positif. Mulailah mengonsumsi 100 mg dan Imromac. Hasilnya, para pasien mulai pulih dan kesehatan mereka mulai membaik.</p> <p>Setelah periode penemuan ilmiah, dokter Rusia menjelaskan bahwa penyakit ini adalah tipuan global, dan metode pengobatan ini menjelaskan, "Ini tidak lain adalah gumpalan di pembuluh darah (bekuan darah) dan metode pengobatan. Tablet antibiotik anti-inflamasi dan Minum antikoagulan (aspirin). Untuk</p>	<p>rusia menjadi negara pertama di dunia yang melakukan otopsi (post mortem) terhadap jenazah covid-19. setelah dilakukan penyelidikan menyeluruh, ditemukan bahwa covid-19 tidak ada dalam bentuk virus, melainkan bakteri yang telah terpapar radiasi dan menggumpal melalui darah hingga menyebabkan kematian. penyakit covid-19 telah ditemukan menyebabkan pembekuan darah manusia dan pembekuan darah vena, yang membuat orang sulit bernapas karena otak, jantung, dan paru-paru tidak dapat menyerap oksigen, menyebabkan orang mati dengan cepat. guna mengetahui penyebab kurangnya energi pernapasan, dokter rusia tidak mendengarkan kesepakatan who, melainkan melakukan otopsi terhadap covid-19. setelah dokter membuka lengan, kaki, dan bagian tubuh lainnya dan memeriksanya dengan cermat, mereka menemukan bahwa pembuluh darah melebar dan berisi gumpalan darah, yang menghalangi aliran darah dan mengurangi aliran oksigen. hal tersebut dapat menyebabkan kematian pada tubuh. setelah mengetahui penelitian tersebut, kementerian kesehatan rusia segera mengubah rencana pengobatan covid-19 dan menggunakan aspirin untuk pasien positif. mulailah mengonsumsi 100 mg dan imromac. hasilnya, para pasien mulai pulih dan kesehatan mereka mulai membaik. setelah periode penemuan ilmiah, dokter rusia menjelaskan bahwa penyakit ini adalah tipuan global, dan metode pengobatan ini menjelaskan, ini tidak lain adalah gumpalan di pembuluh darah (bekuan darah) dan metode pengobatan. tablet antibiotik anti-inflamasi dan minum antikoagulan (aspirin). untuk tujuan ini, kesepakatan telah dikeluarkan di rusia. bagikan</p>

		<p>tujuan ini, kesepakatan telah dikeluarkan di Rusia.</p> <p>Bagikan informasi ini dengan keluarga, tetangga, kenalan, teman, dan kolega Anda sehingga mereka dapat menghilangkan rasa takut akan Covid-19 dan menyadari bahwa itu bukan virus, melainkan bakteri yang hanya terpapar radiasi.</p> <p>Hanya orang dengan kekebalan rendah yang harus berhati-hati. Radiasi ini juga dapat menyebabkan peradangan dan hipoksia. Korban harus mengonsumsi Asprin-100mg dan Apronik atau parasetamol 650mg.</p> <p>Sumber: Kementerian Kesehatan Rusia”</p>	<p>informasi ini dengan keluarga, tetangga, kenalan, teman, dan kolega anda sehingga mereka dapat menghilangkan rasa takut akan covid-19 dan menyadari bahwa itu bukan virus, melainkan bakteri yang hanya terpapar radiasi. hanya orang dengan kekebalan rendah yang harus berhati-hati. radiasi ini juga dapat menyebabkan peradangan dan hipoksia. korban harus mengonsumsi aspirin-100mg dan apronik atau parasetamol 650mg.</p> <p>sumber: kementerian kesehatan rusia</p>
	Non Hoax (FN(1,0))	<p>Catalunya - Barcelona telah menunjuk Menteri Koordinator Bidang Kemaritiman dan Investasi, Luhut Binsar Pandjaitan sebagai pelatih kepala setelah kepergian Ronald Koeman, kata klub Liga Spanyol itu dalam sebuah pernyataan, Kamis (28/10/2021).</p> <p>Koeman dipecat setelah Barca kalah 1-0 dari Rayo Vallecano pada Rabu, kekalahan keempat mereka dalam enam pertandingan di semua kompetisi. "(Luhut) akan mengambil posisi sebagai pelatih kepala tim utama penuh waktu untuk menggantikan Ronald Koeman yang dipecat," kata klub Catalan dalam sebuah pernyataan. Manajemen Barcelona mengatakan, penunjukan Luhut Binsar Pandjaitan sebagai pelatih kepala Barcelona tak lepas dari kinerja apik yang ditunjukkan oleh Luhut di pemerintahan Jokowi-Ma'ruf. Menurut manajemen Barcelona, mereka terkesima melihat performa Luhut ketika ditunjuk menjadi Koordinator PPKM Jawa - Bali, juga ketua Dewan Pengarah Penyelamatan 15 Danau Prioritas Nasional, serta ketua Komite Kereta Cepat Jakarta - Bandung. Hal ini membuat manajemen Barcelona tak ragu untuk menunjuk Luhut sebagai pelatih kepala.</p> <p>*artikel berita berjudul "Barcelona Tunjuk Luhut Panjaitan Sebagai Pelatih Menggantikan Ronald Koeman" Kamis, 28 Oktober 2021 21:51 WIB Oleh: Surya Lesmana/LES</p> <p>=====</p>	<p>catalunya barcelona telah menunjuk menteri koordinator bidang kemaritiman dan investasi, luhut binsar pandjaitan sebagai pelatih kepala setelah kepergian ronald koeman, kata klub liga spanyol itu dalam sebuah pernyataan, Kamis (28 / 10 / 2021). koeman dipecat setelah barca kalah 1 - 0 dari rayo vallecano pada rabu, kekalahan keempat mereka dalam enam pertandingan di semua kompetisi. (luhut) akan mengambil posisi sebagai pelatih kepala tim utama penuh waktu untuk menggantikan ronald koeman yang dipecat, kata klub catalan dalam sebuah pernyataan. manajemen barcelona mengatakan, penunjukan luhut binsar pandjaitan sebagai pelatih kepala barcelona tak lepas dari kinerja apik yang ditunjukkan oleh luhut di pemerintahan jokowi ma ruf. menurut manajemen barcelona, mereka terkesima melihat performa luhut ketika ditunjuk menjadi koordinator ppkm jawa bali, juga ketua dewan pengarah penyelamatan 15 danau prioritas nasional, serta ketua komite kereta cepat jakarta bandung. hal ini membuat manajemen barcelona tak ragu untuk menunjuk luhut sebagai pelatih kepala. * artikel berita berjudul barcelona tunjuk luhut panjaitan sebagai pelatih menggantikan ronald koeman Kamis, 28 oktober 2021 21 : 51 wib oleh : surya lesmana / les</p>
Non Hoax	Hoax (FP(0,1))	<p>Jakarta, CNN Indonesia – Anggota Komisi I DPR RI dari Fraksi PDIP Effendi Simbolon meminta maaf terkait pernyataannya yang menyebut TNI seperti gerombolan. Permintaan maaf itu disampaikan bersama Ketua Fraksi PDIP Utut Adianto di Kompleks Parlemen Senayan, Jakarta, Rabu (14/9).Effendi</p>	<p>permintaan maaf lengkap effendi simbolon sebut tni seperti gerombolan anggota komisi i dpr ri dari fraksi pdip effendi simbolon meminta maaf terkait pernyataannya yang menyebut tni seperti gerombolan. permintaan maaf itu disampaikan bersama ketua fraksi pdip utut adianto di kompleks parlemen</p>

menyinggung TNI seperti gerombolan hingga menuai kecaman dari prajurit. Pernyataannya itu dilontarkan dalam Rapat Dengar Pendapat (RDP) Komisi I DPR RI dengan Kementerian Pertahanan dan TNI di Senayan, Jakarta, Senin (5/9). Lihat Juga : Effendi Simbolon Kirim WhatsApp Minta Maaf: Pak Dudung Belum Merespons

Saat itu, petinggi TNI dari Panglima hingga seluruh kepala staf angkatan hadir, kecuali KSAD Jenderal Dudung Abdurachman. Ketidakhadiran Dudung inilah kemudian menyulut Effendi Simbolon melontarkan kritiknya terhadap TNI. Effendi mengaku ingin mendapat penjelasan dari Panglima TNI Jenderal Andika Perkasa dan KSAD Jenderal Dudung Abdurachman. Menurutnya ada ketidakharmonisan antara dua jenderal bintang empat itu. Berikut permintaan maaf lengkap dari Effendi Simbolon. Saya ingin sampaikan sehubungan dengan di RDP (Rapat Dengar Pendapat) atau Raker Komisi I DPR dengan Kementerian Pertahanan, Panglima, dan para staf. Pokok bahasan di Raker anggaran RKAL 2023 dan isu aktual. Dalam rapat itu kemudian oleh pimpinan disampaikan topik bahasan sesuai undangan. Lihat Juga : Effendi Simbolon Minta Maaf ke Prajurit, KSAD, hingga Panglima TNI Disitu kemudian saya paham ada beberapa poin saya ingin dapat penjelasan dari Menhan, Panglima, Kepala Staf. Memang karena masih pagu indikatif enggak banyak bahasan. Nah, kemudian masuk ke isu aktual. Di situ saya ingin tanya kepada Panglima Andika dan KSAD Dudung, seyogyanya ada Menhan untuk tanyakan info yang kami terima, sehubungan hal-hal disharmoni, menyangkut keberadaan TNI keseluruhan dan AD. Tapi memang Pak KSAD tidak hadir. Oleh teman-teman ditanyakan, dikritisi. Poin saya bukan soal hadirnya tapi lebih elok hadir untuk dapat penjelasan seputar masalah yang kami ingin dapat penjelasan. Di situ kemudian ada masalah kepatuhan, kehormatan TNI yang kami tahu kepatuhan. Di situ saya sadar itu tidak nyaman, tidak elok, dan beberapa pihak mungkin tersinggung atas kata-kata dari saya soal gerombolan dan ormas. Yang sejatinya saya enggak pernah stigma kan gerombolan, tapi kalau enggak ada harmoni itu seperti gerombolan atau ormas. Kalau lihat rekaman utuh itu poin saya. Dari lubuk hati terdalam saya

senayan, jakarta, rabu (14 /9). effendi menyinggung tni seperti gerombolan hingga menuai kecaman dari prajurit. pernyataannya itu dilontarkan dalam rapat dengar pendapat (rdp) komisi i dpr ri dengan kementerian pertahanan dan tni di senayan, jakarta, senin (5 /9). saat itu, petinggi tni dari panglima hingga seluruh kepala staf angkatan hadir, kecuali ksad jenderal dudung abdurachman. ketidakhadiran dudung inilah kemudian menyulut effendi simbolon melontarkan kritiknya terhadap tni. effendi mengaku ingin mendapat penjelasan dari panglima tni jenderal andika perkasa dan ksad jenderal dudung abdurachman. menurutnya ada ketidakharmonisan antara dua jenderal bintang empat itu. berikut permintaan maaf lengkap dari effendi simbolon. saya ingin sampaikan sehubungan dengan di rdp (rapat dengar pendapat) atau raker komisi i dpr dengan kementerian pertahanan, panglima, dan para staf. pokok bahasan di raker anggaran rkal 2023 dan isu aktual. dalam rapat itu kemudian oleh pimpinan disampaikan topik bahasan sesuai undangan. memang karena masih pagu indikatif enggak banyak bahasan. nah, kemudian masuk ke isu aktual. di situ saya ingin tanya kepada panglima andika dan ksad dudung, seyogyanya ada menhan untuk tanyakan info yang kami terima, sehubungan hal - hal disharmoni, menyangkut keberadaan tni keseluruhan dan ad. tapi memang pak ksad tidak hadir. oleh teman - teman ditanyakan, dikritisi. poin saya bukan soal hadirnya tapi lebih elok hadir untuk dapat penjelasan seputar masalah yang kami ingin dapat penjelasan. di situ kemudian ada masalah kepatuhan, kehormatan tni yang kami tahu kepatuhan. di situ saya sadar itu tidak nyaman, tidak elok, dan beberapa pihak mungkin tersinggung atas kata - kata dari saya soal gerombolan dan ormas. yang sejatinya saya enggak pernah stigma kan gerombolan, tapi kalau enggak ada harmoni itu seperti gerombolan atau ormas. kalau lihat rekaman utuh itu poin saya. dari lubuk hati terdalam saya mohon maaf atas perkataan saya yang menyinggung dan menyakiti prajurit siapapun dia perwira, tamtama, dan para pihak yang tidak nyaman atas perkataan yang dinilai lain. saya mohon maaf. saya tujukan pada seluruh prajurit bertugas atau purna dan para pihak yang tidak nyaman dan kepada panglima saya mohon maaf, ksad, ksal, ksau yang mungkin

	<p>mohon maaf atas perkataan saya yang menyinggung dan menyakiti prajurit siapapun dia perwira, tamtama, dan para pihak yang tidak nyaman atas perkataan yang dinilai lain. Saya mohon maaf. Saya tujukan pada seluruh prajurit bertugas atau purna dan para pihak yang tidak nyaman dan kepada Panglima saya mohon maaf, KSAD, KSAL, KSAU yang mungkin merasa kurang nyaman saya mohon maaf. Saya ingin menegakkan mencintai TNI dengan tupoksi saya. Tidak kurangi hormat kita ke hal-hal kurang pas. Demikian. Lihat Juga :Dandim Cilegon Murka Gebrak Meja, Suruh Effendi Simbolon Minta Maaf Apakah Bapak sudah minta maaf langsung kepada Panglima dan KSAD? Dua hari lalu saya WhatsApp Panglima dan KSAD. Kemarin saya ketemu Panglima menanyakan sikap TNI dan menyampaikan maaf saya. Panglima sampaikan tidak ada masalah jadi sangat klir. Silakan teman-teman tanya langsung lebih elok yang bersangkutan yang menyampaikan. Dudung belum merespons. Saya bertanggung jawab apa yang saya sampaikan. Misi saya datang untuk minta maaf dan saya sudah sampaikan ke Panglima, ke KSAD belum. Saya bukan pemilik kebenaran, saya hanya ingin mendapatkan tentang disharmoni yang mengganggu. Demi Allah Tuhan saya tidak pernah men-judge seperti yang beredar, apalagi saya bagian dari keluarga besar TNI. (cfd/pmg)</p>	<p>merasa kurang nyaman saya mohon maaf. saya ingin menegakkan mencintai tni dengan tupoksi saya. tidak kurangi hormat kita ke hal - hal kurang pas. demikian. kemarin saya</p>
<p>Non Hoax (TN(0,0))</p>	<p>Jakarta, CNN Indonesia -- Ketua Komisi I DPR RI, Meutya Hafid, mengungkapkan bahwa waktu penyelenggaraan uji kelayakan dan kepatutan (fit and proper test) calon Panglima TNI Jenderal Andika Perkasa baru diputuskan dalam rapat internal komisinya yang digelar pada Kamis (4/11).Menurutnya, Komisi I DPR belum memutuskan waktu penyelenggaraan uji kelayakan dan kepatutan Andika hingga saat ini. Komisi I akan mengadakan rapat internal esok siang [Kamis]pukul 14.00 WIB. Rapat internal yang diikuti seluruh anggota komisi 1 akan memutuskan kapan fit and proper test diadakan," kata Meutya kepada wartawan, Rabu (3/11).</p> <p>Dia pun menerangkan lazimnya proses pemilihan pemilihan Panglima TNI, Komisi I DPR diberikan waktu untuk melakukan verifikasi dokumen hingga aktual dengan mengunjungi kediaman calon Panglima TNI.Lihat Juga : Jokowi</p>	<p>waktu fit and proper test andika jadi panglima ditentukan besok ketua komisi i dpr ri, meutya hafid, mengungkapkan bahwa waktu penyelenggaraan uji kelayakan dan kepatutan (fit and proper test) calon panglima tni jenderal andika perkasa baru diputuskan dalam rapat internal komisinya yang digelar pada kamis (4 / 11). menurutnya, komisi i dpr belum memutuskan waktu penyelenggaraan uji kelayakan dan kepatutan andika hingga saat ini. "" komisi i akan mengadakan rapat internal esok siang [kamis] pukul 14. 00 wib. rapat internal yang diikuti seluruh anggota komisi 1 akan memutuskan kapan fit and proper test diadakan, "" kata meutya kepada wartawan, rabu (3 / 11). dia pun menerangkan lazimnya proses pemilihan pemilihan panglima tni, komisi i dpr diberikan waktu untuk melakukan verifikasi dokumen hingga aktual dengan mengunjungi kediaman calon panglima tni. namun, ia berkata proses itu harus</p>

	<p>Tunjuk Andika Panglima TNI, Pakar Sorot Faktor Hendropriyono Meutya memastikan, Komisi I DPR ingin memproses pencalonan Andika berjalan cepat. Namun, ia berkata proses itu harus dilakukan dengan memperhatikan semua tahapan yang perlu dilalui secara lengkap. Meutya menambahkan, komisinya masih menunggu penugasan dari Badan Musyawarah (Bamus) DPR untuk memproses pencalonan Andika sebagai Panglima TNI."Saat ini Komisi I masih menunggu surat penugasan. Hingga lewat pukul 18.00 hari ini, Komisi I DPR belum menerima surat penugasan dari Bamus. Jadi kami masih menunggu," ujar Ketua DPP Golkar itu. Lihat Juga : Golkar Sebut Pangkostrad Dudung Calon KSAD Pengganti AndikaSebelumnya, Ketua DPR, Puan Maharani, mengatakan Komisi I DPR akan menggelar uji kelayakan dan kepatutan Andika mulai besok Kamis (4/11) hingga Jumat (5/11). November 2021.Menurutnya, pihaknya akan menggelar rapat Bamus untuk membahas Surat Presiden (Surpres) soal pencalonan Andika serta menjadwalkan waktu penyelenggaraan uji kelayakan dan kepatutan calon Panglima TNI pada hari ini. "Begitu Surpres diterima hari ini, Badan Musyawarah DPR langsung menggelar rapat dan menjadwalkan fit and proper test calon Panglima TNI besok sampai dengan Jumat," kata Puan kepada wartawan, Rabu (3/11). Lihat Juga: Pengamat soal Bakal PR Andika: Alutsista, China, dan Tumpukan Perwira (mts/DAL)</p>	<p>dilakukan dengan memperhatikan semua tahapan yang perlu dilalui secara lengkap. meutya menambahkan, komisinya masih menunggu penugasan dari badan musyawarah (bamus) dpr untuk memproses pencalonan andika sebagai panglima tni. "" saat ini komisi i masih menunggu surat penugasan. hingga lewat pukul 18. 00 hari ini, komisi i dpr belum menerima surat penugasan dari bamus. jadi kami masih menunggu, "" ujar ketua dpp golkar itu. november 2021. menurutnya, pihaknya akan menggelar rapat bamus untuk membahas surat presiden (surpres) soal pencalonan andika serta menjadwalkan waktu penyelenggaraan uji kelayakan dan kepatutan calon panglima tni pada hari ini. "" begitu surpres diterima hari ini, badan musyawarah dpr langsung menggelar rapat dan menjadwalkan fit and proper test calon panglima tni besok sampai dengan jumat, "" kata puan kepada wartawan, rabu (3 / 11). mts / dal)</p>
--	---	--

Berdasarkan Tabel 2. dan tabel 3. terlihat bahwa adanya perbedaan panjang narasi antaraberita *hoax* dan *non-hoax*. Berita *hoax* cenderung memiliki narasi yang lebih pendek dibandingkan dengan narasi berita *non-hoax* sehingga model cenderung memprediksi berita sebagai berita *hoax* apabila berita tersebut memiliki narasi yang pendek. Terdapat Beberapa Karakteristik dalam mengetahui berita *hoax*, Karakteristik berita *hoax* sebagai berikut:

- 1) Memiliki Narasi Berita Yang Lebih Pendek atau singkat
- 2) Bahasa penulisan berita yang digunakan cenderung tidak baku dan menysasar topik-topik sensitif seperti politik, agama, ras, atau kesehatan.
- 3) Sering Mencantumkan konten yang menyesatkan atau tidak akurat

Tabel 3. Contoh Berita Hoax

biang hoax kerja sama dengan turn back hoax pantesan komunitas turn back hoax tutup mata sama hoax kompas, ternyata satu afiliasi. penjelasan & fakta : sebenarnya, klarifikasi ini sudah dibuat bulan desember 2016 tepatnya tanggal 24 setelah banyak yang mempertanyakan kenapa ada tulisan kompas. com di poster tersebut. namun karena ada lagi netizen yang kembali membahas soal poster tersebut, maka dengan senang hati admin mengklarifikasi ulang penjelasannya bisa dilihat di sini

IV. KESIMPULAN

Dalam studi ini, total berita yang dianalisis berjumlah 4.391, dengan 3.351 di antaranya merupakan berita hoax dan 1.040 sisanya adalah berita non-hoax. Untuk mencapai model LSTM yang efisien, pelatihan dilakukan dengan *hyperparameter* yang telah dioptimalkan, mencakup 16 neuron LSTM, ukuran *embedding* 512, dan batch size sebanyak 16. Panjang narasi dari berita juga berperan dalam kemampuan prediksi model. Model paling efektif dicapai dengan akurasi sebesar 99,72%, *train loss* hanya 0,15%, *Validation Loss* sebesar 0,02%, menggunakan *learning rate* sebesar $2e-5$, dan melalui 5 *epoch* pelatihan.

REFERENSI

- [1] P. KOMINFO, "Ada 800.000 Situs Penyebar Hoax di Indonesia," Website Resmi Kementerian Komunikasi dan Informatika RI. 2017. [Online].http://content/detail/12008/ada-800000-situs-penyebarhoaxdiindonesia/0/sorotan_media%0Ahttp://files/66/sorotan_media.html.
- [2] I. R. Cahyadi, "Survei KIC: Hampir 60% Orang Indonesia Terpapar Hoax Saat Mengakses Internet," Berita Satu, 2020. <https://www.beritasatu.com/digital/700917/survei-kic-hampir-60-orang-indonesia-terpapar-hoax-saat-mengakses-internet> (accessed Nov. 04,2023).
- [3] Kaur, S., Kumar, P., & Kumaraguru, P. (2020). Detecting clickbaits using two-phase hybrid CNN-LSTM biterm model. *Expert Systems with Applications*, 151. <https://doi.org/10.1016/j.eswa.2020.113350>
- [4] Goonathilake, M. D. P. P., & Kumaral, P. P. N. V. (2020). CNN, RNN-LSTM Based Hybrid Approach to Detect State-of-the-Art Stance-Based Fake News on Social Media. *20th International Conference on Advances in ICT for Emerging Regions, ICTer 2020 - Proceedings*, 23–28. <https://doi.org/10.1109/ICTer51097.2020.9325477>
- [5] Yunanto, R., Purfini, A. P., & Prabuwisesa, A. (n.d.). *Jurnal Manajemen Informatika (JAMIKA) Survei Literatur: Deteksi Berita Palsu Menggunakan Pendekatan Deep Learning*. <https://doi.org/10.34010/jamika.v11i2.493>
- [6] Arbain, A., Muhammad, M. A., Septiana, T., & Septama, H. D. (2022). LEARNING HOAX NEWS PADA LOCAL DAN CLOUD COMPUTING DEPLOYMENT MENGGUNAKAN GOOGLE APP ENGINE. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3). <https://doi.org/10.23960/jitet.v10i3.2646>
- [7] Kurniawan, A. A., & Mustikasari, M. (2020). *Implementasi Deep Learning Menggunakan Metode CNN dan LSTM untuk Menentukan Berita Palsu dalam Bahasa Indonesia*. 5(4), 2622–4615. <https://doi.org/10.32493/informatika.v5i4.7760>
- [8] Isa, S. M., Nico, G., & Permana, M. (2022). INDOBERT FOR INDONESIAN FAKE NEWS DETECTION. *ICIC Express Letters*, 16(3), 289–297. <https://doi.org/10.24507/icicel.16.03.289>
- [9] Utama, L. B., & Suhartono, D. (2022). Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic. *Informatika (Slovenia)*, 46(8), 81–90. <https://doi.org/10.31449/inf.v46i8.4336>
- [10] Goonathilake, M. D. P. P., & Kumaral, P. P. N. V. (2020). CNN, RNN-LSTM Based Hybrid Approach to Detect State-of-the-Art Stance-Based Fake News on Social Media. *20th International Conference on Advances in ICT for Emerging Regions, ICTer 2020 - Proceedings*, 23–28. <https://doi.org/10.1109/ICTer51097.2020.9325477>.