

Optimalisasi Summarization Berita BBC dengan Metode BiLSTM-Transformer

Rafael Austin ¹, Andhika Dwi Rachmawanto ^{2*}, Michael Jeconiah Yonathan ³, M Naufal Arriz⁴, dan Vitri Tundjungsari ⁵

¹ Universitas Esa Unggul, Jalan Arjuna Utara 9, Jakarta, Jawa Barat;
e-mail : raf.austin05@student.esaunggul.ac.id

² Universitas Esa Unggul, Jalan Arjuna Utara 9, Jakarta, Jawa Barat;
e-mail : dwirahmawanto39@student.esaunggul.ac.id

³ Universitas Esa Unggul, Jalan Arjuna Utara 9, Jakarta, Jawa Barat;
e-mail : michaeljeconiah09@student.esaunggul.ac.id

⁴ Universitas Esa Unggul, Jalan Arjuna Utara 9, Jakarta, Jawa Barat;
e-mail : arriznaufal@student.esaunggul.ac.id

⁵ Universitas Esa Unggul, Jalan Arjuna Utara 9, Jakarta, Jawa Barat;
e-mail : vitri.tundjungsari@esaunggul.ac.id

* Corresponding Author : Andhika Dwi Rachmawanto

Abstract: The rapid growth of digital news, such as that from the BBC, presents challenges for readers in absorbing dense information within limited time. This research proposes an automated text summarization system using a hybrid BiLSTM Transformer architecture to produce concise yet contextually accurate summaries. The model integrates BiLSTM to capture local sequential relationships and Transformer's self-attention mechanism to handle global context, overcoming the computational limitations of standalone Transformers. Utilizing a self-embedding approach, the system processes text in an unsupervised manner, making it suitable for datasets without ground truth summaries. Evaluation was conducted using 50 samples from the Xsum dataset and 25 live BBC news links, with performance measured via cosine similarity to assess contextual preservation. The results demonstrated a consistent average cosine similarity of 0.7959 for dataset samples and 0.7877 for new data. These findings indicate that the hybrid model effectively maintains semantic integrity and provides reliable summaries for complex news articles.

Keywords : BiLSTM; AI; Transformer; Embedding; Summarizer.

Abstrak: Pesatnya pertumbuhan berita digital, seperti dari media BBC, menghadirkan tantangan bagi pembaca dalam menyerap informasi yang padat dalam waktu terbatas. Penelitian ini mengusulkan sistem peringkasan teks otomatis menggunakan arsitektur hibrida BiLSTM Transformer untuk menghasilkan ringkasan yang ringkas namun akurat secara kontekstual. Model ini mengintegrasikan BiLSTM untuk menangkap hubungan sekuensial lokal dan mekanisme self-attention dari Transformer untuk menangani konteks global, guna mengatasi keterbatasan komputasi pada model Transformer mandiri. Dengan menggunakan pendekatan self-embedding, sistem memproses teks secara unsupervised, sehingga sesuai untuk dataset yang tidak memiliki ringkasan acuan (ground truth). Evaluasi dilakukan menggunakan 50 sampel dari dataset Xsum dan 25 tautan berita BBC langsung, dengan kinerja yang diukur melalui cosine similarity untuk menilai pemeliharaan konteks. Hasil penelitian menunjukkan rata-rata cosine similarity yang konsisten sebesar 0,7959 untuk sampel dataset dan 0,7877 untuk data baru. Temuan ini mengindikasikan bahwa model hibrida tersebut efektif dalam menjaga integritas semantik dan memberikan ringkasan yang andal untuk artikel berita yang kompleks.

Kata kunci : BiLSTM; AI; Transformer; Embedding; Summarizer.

Naskah Masuk: 24 Januari 2026
Revisi: 30 Januari 2026
Diterima: 16 Februari 2026
Terbit: 9 Maret 2026
Ver. Skrg.: 9 Maret 2026



Copyright: © 2026 by the authors.
Submitted for possible open
access publication under the
terms and conditions of the
Creative Commons Attribution
(CC BY SA) license
(<https://creativecommons.org/licenses/by-sa/4.0/>)

1. Pendahuluan

Berita daring dari media internasional seperti BBC umumnya disajikan dalam teks panjang dan padat informasi untuk memberikan konteks yang komprehensif. Hal ini membuat pembaca membutuhkan waktu lebih lama untuk memahami inti informasi [1]. Seiring dengan meningkatnya jumlah berita digital, pembaca dihadapkan pada keterbatasan waktu dan perhatian dalam menyerap keseluruhan isi berita [1]. Oleh karena itu, diperlukan metode peringkasan teks otomatis untuk menyajikan informasi inti secara ringkas tanpa menghilangkan makna utama [1]. Namun, dalam melakukan peringkasan teks, diperlukan kemampuan untuk memahami konteks serta menghindari adanya redundansi informasi [2].

Embedding adalah teknik representasi data yang memetakan kata, kalimat, atau dokumen ke dalam vektor berdimensi rendah agar hubungan semantik dapat direpresentasikan secara numerik. Self embedding adalah pendekatan yang mempelajari representasi vektor dari struktur internal data tanpa memerlukan label eksplisit dan termasuk dalam unsupervised atau self-supervised learning. Pendekatan self embedding digunakan karena data tidak memiliki ground truth, sehingga representasi konteks diekstraksi dari pola distribusional data [3]. Pada penelitian sebelumnya, penggunaan cosine similarity sebagai pengukuran konteks dalam teks terbukti efektif dan memberikan hasil yang baik [4]. Oleh karena itu, pada penelitian ini cosine similarity digunakan sebagai metrik untuk mengukur kinerja sistem peringkasan dalam menjaga kesesuaian konteks antara teks asli dan teks hasil ringkasan.

BiLSTM adalah pengembangan LSTM yang memproses data sekuensial dari dua arah, yaitu masa lalu dan masa depan. Model Transformer dua arah seperti BERT memungkinkan penyematan kontekstual yang lebih kaya dengan mengkondisikan secara bersamaan pada konteks kiri dan kanan di semua lapisan [5]. Arsitektur Transformer teks-ke-teks terpadu menunjukkan kinerja yang kuat di berbagai tugas NLP, termasuk peringkasan, dengan memanfaatkan pembelajaran transfer dari pelatihan awal skala besar [6]. Jaringan generator penunjuk menggabungkan mekanisme penyalinan dan model sekuens-ke-sekuens untuk mengatasi masalah di luar kosakata dan meningkatkan konsistensi faktual dalam peringkasan abstrak [7]. Salah satu varian RNN yang paling dikenal adalah Long Short-Term Memory (LSTM) [8]. Bidirectional Long Short-Term Memory (BiLSTM) merupakan pengembangan dari LSTM yang dirancang untuk memproses data sekuensial dengan mempertimbangkan konteks informasi dari dua arah, yaitu masa lalu dan masa depan [9]. Variasi ini juga mengurangi salah satu kelemahan LSTM yaitu vanishing gradient dengan memperkaya aliran informasi melalui gradien dua arah [10]. Lapisan BiLSTM mampu menangkap struktur internal kalimat serta peran kontekstual dalam dokumen, terutama pada kerangka representasi hierarkis.

Sebelum berkembangnya Transformer, BiLSTM merupakan model yang dominan digunakan dalam berbagai tugas pemrosesan teks [11]. Namun, penggunaan BiLSTM kemudian bergeser ke Transformer karena adanya mekanisme self-attention yang mampu menangkap hubungan global antar kata secara lebih efektif [12]. Transformer adalah arsitektur jaringan neural yang sepenuhnya menggunakan mekanisme attention, tanpa rekuren maupun konvolusi, untuk menangani tugas pemrosesan urutan secara paralel dan efisien [13]. Meskipun demikian, Transformer memiliki ketergantungan terhadap kebutuhan daya komputasi yang tinggi, sehingga penggunaan dataset berukuran besar dapat menjadi kendala [12]. Transformer merupakan metode representasi tekstual yang menggunakan mekanisme self-attention untuk menangkap ketergantungan global antar elemen dalam data sekuensial dan spasial secara efisien. Dalam konteks peringkasan teks, mekanisme ini bekerja antar kata untuk memahami konteks secara menyeluruh [11]. Namun, untuk mengatasi keterbatasan komputasi sekaligus mempertahankan kemampuan pemodelan konteks lokal dan global, beberapa penelitian sebelumnya melakukan penggabungan arsitektur BiLSTM dan Transformer [11].

Evaluasi arsitektur LSTM dan Transformer untuk text summarization menunjukkan bahwa model Transformer menghasilkan ringkasan yang lebih akurat dan ringkas dibandingkan LSTM dalam tugas peringkasan otomatis. Pada penelitian ini di tambah lagi dengan ada nya bidirectional [14]. Analisis komparatif antara arsitektur Transformer dan LSTM pada task text summarization menunjukkan bahwa Transformer unggul dalam

menangani dokumen panjang dengan skor ROUGE yang lebih tinggi dibandingkan LSTM [15]. Penelitian ini menggunakan perbedaan dengan menggunakan BiLSTM dan transformer. Kedua teknik tersebut terbukti efektif dalam menangani pemodelan sekuensial dalam pemahaman konteks global pada penelitian sebelumnya, sehingga dikombinasikan dalam arsitektur BiLSTM Transformer untuk memanfaatkan keunggulan masing-masing metode, yaitu pemahaman konteks secara lokal dan global [16]. Berbeda dengan pendekatan lain seperti sentence transformer yang digunakan pada sistem rekomendasi dokumen [17], pendekatan BiLSTM Transformer lebih sesuai untuk sistem peringkasan karena mampu mempelajari struktur kalimat dan melakukan proses sequence-to-sequence [17]. Penelitian ini mengusulkan arsitektur BiLSTM Transformer sebagai pendekatan hibrida untuk menghasilkan ringkasan yang informatif, akurat, dan tetap menjaga konteks semantik. menjaga kesesuaian konteks semantik secara optimal.

2. Kajian Pustaka

Penelitian mengenai peringkasan teks otomatis (automatic text summarization) berkembang pesat seiring meningkatnya jumlah berita digital yang panjang dan padat informasi. Metode awal peringkasan umumnya berbasis statistik, seperti frekuensi kata dan pemilihan kalimat penting, namun pendekatan tersebut memiliki keterbatasan dalam memahami konteks semantik dan hubungan antar kalimat. Oleh karena itu, metode ini sering menghasilkan ringkasan yang kurang koheren dan tidak sepenuhnya merepresentasikan makna utama dokumen.

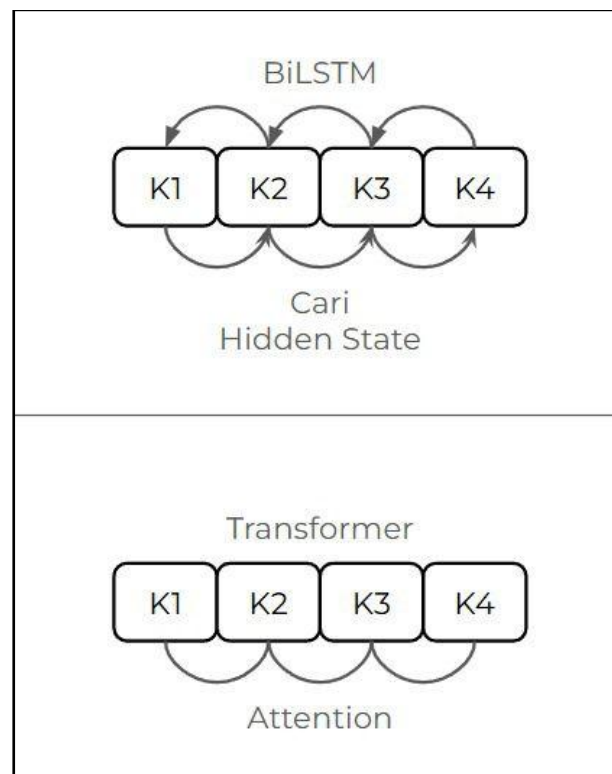
Perkembangan deep learning mendorong penggunaan jaringan saraf berbasis sekuensial, khususnya Long Short-Term Memory (LSTM) dan Bidirectional LSTM (BiLSTM). BiLSTM mampu memproses informasi dari dua arah sehingga memberikan representasi kontekstual yang lebih kaya dibandingkan LSTM satu arah. Dalam penelitian sebelumnya, BiLSTM terbukti efektif dalam menangkap struktur kalimat dan hubungan lokal antar kata, sehingga banyak digunakan dalam tugas pemrosesan bahasa alami, termasuk peringkasan teks berita.

Selanjutnya, diperkenalkan arsitektur Transformer yang mengandalkan mekanisme self-attention untuk menangkap hubungan global antar kata dalam dokumen secara efisien dan paralel. Transformer menunjukkan kinerja yang sangat baik dalam memahami konteks global, namun memiliki kelemahan pada kebutuhan komputasi yang tinggi serta kurang optimal dalam menangani dependensi lokal secara eksplisit. Kondisi ini mendorong munculnya pendekatan hibrida yang mengombinasikan keunggulan BiLSTM dan Transformer.

Beberapa penelitian terdahulu menunjukkan bahwa arsitektur BiLSTM Transformer mampu menghasilkan representasi teks yang lebih informatif dengan menggabungkan pemahaman konteks lokal dan global secara seimbang. Selain itu, penggunaan pendekatan unsupervised berbasis embedding memungkinkan sistem peringkasan bekerja tanpa ground truth. Evaluasi menggunakan cosine similarity banyak digunakan untuk mengukur kesesuaian konteks antara teks asli dan ringkasan. Berdasarkan kajian tersebut, penelitian ini memanfaatkan arsitektur BiLSTM Transformer untuk menghasilkan ringkasan berita BBC yang ringkas, kontekstual, dan semantik tetap terjaga.

3. Metode yang Diusulkan

Model hybrid BiLSTM Transformer dirancang untuk memanfaatkan kemampuan BiLSTM dalam menangkap hubungan lokal dalam urutan sekaligus kemampuan Transformer dalam menangkap dependensi konteks global. Penelitian ini menggunakan 1 lapisan encode yaitu pada kalimat [18] karena fokus kepada teks itu sendiri dan bukan dalam bentuk dokumen



Gambar 1. Representasi cara kerja BiLSTM dan Transformer

Cara kerja dan perbedaan kedua algoritma ini dapat ditunjukkan pada gambar 1 dimana untuk BiLSTM, algoritma akan mengecek satu kata secara berurutan dari kiri dan kanan seiring mengumpulkan hidden state dari setiap dua kata yang dicek secara berurutan sementara Transformer bekerja dengan cara mencari konteks global setiap kalimat seperti pada gambar 1 dimana K3 akan membentuk attention pada K lainnya seperti K1, K2, dan K4. Kedua parameter ini akan membantu dalam proses melakukan summarization pada teks.

Integrasi BiLSTM dan Transformer dilakukan dengan metode stacking, di mana output dari BiLSTM menjadi input langsung bagi Transformer Encoder. Karena BiLSTM bersifat bidirectional, hidden state dari arah maju dan mundur digabung menggunakan mekanisme concatenation, sehingga menghasilkan representasi berdimensi $2 \times \text{hidden_dim}$ untuk setiap token. Bentuk tensor keluarannya dapat dinyatakan sebagai $(\text{batch_size}, \text{seq_len}, 2 \times \text{hidden_dim})$, yang kemudian disesuaikan agar sama dengan d_model pada Transformer. Di dalam Transformer, representasi tersebut diproyeksikan menjadi Query, Key, dan Value dengan dimensi d_model untuk menjalankan mekanisme self-attention. Output akhir tetap berbentuk $(\text{batch_size}, \text{seq_len}, d_model)$ sebelum dipetakan ke ruang kosakata melalui linear layer.

Output dari proses tersebut adalah representasi kontekstual global untuk setiap token yang telah mengandung informasi dua arah dari BiLSTM serta pembobotan relevansi antar token dari mekanisme self-attention. Representasi ini kemudian diproyeksikan melalui linear layer menjadi skor pada ruang kosakata dengan dimensi $(\text{batch_size}, \text{seq_len}, \text{vocab_size})$. Setiap nilai pada dimensi terakhir merepresentasikan probabilitas relatif suatu token terhadap seluruh kosakata. Setelah itu, fungsi aktivasi seperti softmax digunakan untuk menghasilkan distribusi probabilitas akhir. Token dengan probabilitas tertinggi dipilih sebagai prediksi keluaran model pada setiap posisi sekuens.

Model ini akan diimplementasikan pada program summarizer dengan urutan dilakukannya BiLSTM dahulu dilanjutkan dengan penggunaan Transformer pada akhir program. Model hybrid BiLSTM Transformer dirancang untuk memanfaatkan kemampuan BiLSTM dalam menangkap hubungan lokal dalam urutan sekaligus kemampuan Transformer dalam menangkap dependensi konteks global [19].

Dataset yang digunakan pada penelitian ini berasal dari web terbuka huggingface. Data yang dipakai bernama Xsum yang berisi text berita BBC secara langsung dalam bentuk file. Dataset ini berbentuk file satuan yang berisi link, judul berita, serta seluruh text berita itu sendiri. Jumlah file yang terdapat dalam dataset ini adalah 237.018 file. Jumlah ini sangat banyak dan dapat memakan waktu yang sangat banyak karena itu dalam penelitian ini digunakan sebanyak 500 file. Program tidak memakai ground truth sehingga menggunakan pendekatan tanpa supervisi eksplisit digunakan dalam proses peringkasan teks secara langsung dalam program.

Berikut ini adalah contoh isi dari salah satu file yang terdapat pada dataset xsum

```
[SH]URL[SH]
http://web.archive.org/web/20110610181025/http://www.bbc.co.uk/newsbeat/10000983

[SH]TITLE[SH]
Anger over 'US criticism' of NHS

[SH]FIRST-SENTENCE[SH]
It's being called "evil" and a "death panel" where bureaucrats decide who lives and who dies. Any ideas what it could be?

[SH]TEXTBODY[SH]
Well, it's how the idea of Britain's NHS is being described over in America.
It's all part of a backlash against Barack Obama's planned changes to the US health system.
There's been an angry response from Newsbeat listeners who point out the UK is above the US in healthcare league tables.
Fiona from Derby says the NHS saved her life. "I almost died of Hodgkin's lymphoma aged 20," she said. "I've since gone on to have two children. I'm owed their lives as much as my own."
Gareth in Hampshire has had a personal experience of the health system in the US. "I lived in the USA and the healthcare system is designed for the rich," he texted.
"my two-year-old son had to go to the emergency room for a high fever.
"The hospital charged $10,000 (£6,076). He only had three injections and a quick check up."
John is a medical student at the University of Manchester. He thinks the adverts are just scare tactics. "The American adverts about our health service are just wrong," he said.
"The health service in the UK is scored way higher than that of the US. You just need to watch the Micheal Moore film Sicko to see that those adverts are hollow propaganda."
Mike in Dundee thinks these are American fear tactics. He said: "They're all so afraid of losing their own money they resort to attacking their only ally left in the world."
But Andrew from Rochester in Kent agrees with Sarah Palin.
He said: "The NHS should be scrapped, why should hard working people be taxed to death to pay for healthcare for those who can't even be bothered to get out of bed and then have to pay even more for private treatment because the government has wasted all the tax income?"
This person didn't leave their name but says they've been ill for the past 11 months.
"I have seen countless doctors and been given loads of tablets. Nothing has helped but they refuse to test me because I'm not a high-risk patient because I'm not overweight.
"The last doctor I saw said they didn't know what it was and I'll probably have it for the next 10 years!
"The NHS just want to fob people off with tablets rather than actually care for their illnesses."
```

Gambar 2. Potongan Gambar file sampel dari dataset

Setiap File terdiri dari 3 bagian utama yang menunjukkan beberapa informasi. URL akan memberikan link yang akan membawa ke link asli pada web berita BBC secara langsung. Link menggunakan fungsi wayback machine pada Web Berita BBC untuk mencari berita lama yang ada dalam dataset. Untuk bagian kedua berisi judul berita dari Web BBC secara langsung. Bagian ketiga berisi isi dari berita tersebut yaitu bagian yang akan dipakai dalam melakukan embedding dan training pada algoritma BiLSTM-Transformer Hybrid

Pada penelitian ini digunakan pendekatan self embedding sebagai teknik representasi teks, di mana kalimat dipetakan ke dalam bentuk vektor numerik yang dipelajari langsung dari struktur internal data tanpa memerlukan ground truth. Pendekatan ini dipilih karena dataset yang digunakan tidak menyediakan ringkasan acuan yang tervalidasi, sehingga pembentukan representasi konteks dilakukan secara unsupervised untuk mendukung proses peringkasan ekstraktif [20]. Dengan menggunakan self embedding, model mampu menangkap hubungan semantik antar kata dan frasa di dalam kalimat. Representasi numerik yang dihasilkan mempermudah pemrosesan oleh model BiLSTM dan Transformer. Selain itu, pendekatan ini memungkinkan fleksibilitas dalam menangani variasi panjang kalimat dan kompleksitas struktur teks.

Evaluasi pada penelitian ini dilakukan dengan cara melakukan perbandingan antara teks sampel sebelum dilakukan summarization dan sesudah summarization. Evaluasi ini dilakukan dengan menggunakan cosine similarity untuk mengukur kemampuan peringkasan dalam menjaga konteks. Evaluasi akan dilakukan sebanyak 2 kali. Yang pertama akan dilakukan dengan cara memakai 50 data sampel dari dataset dan dilanjutkan dengan evaluasi dengan menggunakan 25 link data secara langsung dari web bbc untuk mengukur kinerja model. Berikut ini adalah rumus Cosine Similarity yang ditunjukkan oleh persamaan (1)

$$\text{sim}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

Program akan akan dibentuk dan dibuat pada google collab dan menggunakan bahasa pemrograman python. Untuk dataset akan diupload secara langsung pada google drive dan bisa langsung diakses dengan link pada google collab itu sendiri. program akan memakai beberapa library yaitu

Tabel 1. Library dan Kegunaannya

Library	Kegunaan
Gdown	Untuk pengambilan file dari google drive link secara langsung dan langsung di unduh pada lingkungan google collab dengan penggunaan link seperti

	contoh berikut adalah link secara langsung dari drive yang akan secara langsung juga di unduh pada lingkungan google collab. https://drive.google.com/file/d/1GLWLRGZo3z0TYfXQY35lOzM0lSvKZl--/view?usp=sharing
Tarfile	Untuk Ekstraksi tipe file tar.gz menjadi pecahan file aslinya yang akan dikelompokkan oleh library os lagi.
Os	Untuk melakukan navigasi pada file. Pada program libraru ini akan digunakan untuk melakukan pengambilan sample sebanyak 500 file untuk dipakai sebagai dataset
torch	library ini adalah framework utama yang memungkinkan BiLSTM melakukan memproses data sekuensial seperti teks yang dipakai untuk dataset sementara untuk untuk transformer, library ini berisi komposisi pembentuk dari algoritma transformer yaitu MultiheadAttention, LayerNorm, serta Embedding itu sendiri. Library ini memungkinkan dilakukan nya proses hybrid antar kedua algoritma
request	Library ini digunakan pada proses evaluasi dengan melakukan request dengan link bbc terkini dan melakukan pengambilan data teks berita dalam web itu sendiri
sklearn	Library ini berisi fungsi pra pemrosesan pada data seperti pembagian dataset, encoding, hingga pada akhir evaluasi kinerja program yaitu cosine similarity yang bisa mengukur kemiripan dua objek
Numpy	Library ini berisi fungsi manipulasi pada array, perhitungan matriks, serta reshaping pada data agar proses BiLSTM Transformer dapat dilakukan pada dataset
matplotlib	Library ini berisi fungsi yang dapat dipakai sebagai visualisasi hasil dari program dengan cara membuat grafik grafik yang dapat dengan mudah menunjukan efesiensi dan kinerja dari algoritma pada perbandingan akhirnya

Prosedur penelitian ini terdiri dari 3 proses yaitu Training data hingga testing dengan menggunakan salah satu sampel pada dataset untuk menunjukan program bekerja lalu Evaluasi kinerja program dalam menangkap konteks dari teks panjang artikel dengan menggunakan 25 data sampel dataset dan akhiri dengan penggunaan 25 link artikel bbc untuk membandingkan kemampuan model dalam melakukan summarizing pada berita bbc baru diluar dari dataset itu sendiri

4. Hasil dan Pembahasan

Pada bagian ini akan dipresentasikan hasil dari program sesuai dari prosedur yang diberikan pada metode yaitu pembuatan program, testing awal, evaluasi dataset, dan testing evaluasi dengan link baru bbc. Setiap program dalam penelitian ini dijalankan dan dibuat dalam google collab secara langsung. Setiap Evaluasi akan menggunakan cosine similarity untuk menilai kinerja model dalam menjaga konteks hasil summarizer.

4.1. Implementasi Awal Model

Tahap ini dimulai dengan melakukan pengunduhan dataset dari drive dengan library gdown dan disimpan pada lingkungan google collab dan bisa dipakai. Setelah dilakukannya pengunduhan dilanjutkan pada ekstraksi dengan library Tarfile karena file yang diunduh terdapat dalam bentuk tar.gz atau format kompresi umum yang terdapat pada linux

```

*** Downloading...
From (original): https://drive.google.com/uc?id=1GLWLRGZo3z0TYfXQY35lOzM0lSvKZl--
From (redirected): https://drive.google.com/uc?id=1GLWLRGZo3z0TYfXQY35lOzM0lSvKZl--&confirm=t&uuiid=5
To: /content/xsum.tar.gz
100%|██████████| 255M/255M [00:05<00:00, 50.1MB/s]
/tmp/ipython-input-1634804913.py:16: DeprecationWarning: Python 3.14 will, by default, filter extrac
tar.extractall(path=extract_dir)
✓ Extract selesai ke: /content/xsum_data
i File tersedia di folder: ['bbc-summary-data']

```

Gambar 3. Potongan output program dalam melakukan pengunduhan program

```
Total file tersedia: 0
✅ Subset 500 file disimpan di: /content/xsum_subset
```

Gambar 4. Potongan output program dalam memisahkan sampel

file diekstrak dan masuk kedalam folder yang sudah dibuat oleh library Os yang membantu navigasi file. Setelah dijalankannya ini file file akan bisa ditemukan pada path yang sudah ditentukan dan tersimpan pada lingkungan google collab

```
Parsing files: 100%|██████████| 500/500 [00:00<00:00, 8004.67it/s]Loaded samples: 500

START TRAINING

Epoch 1/3: 100%|██████████| 63/63 [03:26<00:00, 3.27s/it, loss=7.86]
Epoch 1 | Avg Loss: 8.1933
Epoch 2/3: 100%|██████████| 63/63 [03:20<00:00, 3.18s/it, loss=6.88]
Epoch 2 | Avg Loss: 7.0839
Epoch 3/3: 100%|██████████| 63/63 [03:20<00:00, 3.18s/it, loss=7.33]Epoch 3 | Avg Loss: 6.8608

Training selesai. Model tersimpan sebagai 'bilstm_transformer_model.pt'.
```

Gambar 5. Potongan output program dalam melakukan training

Pada program metode loss function yang dipakai adalah Cross-Entropy Loss, yaitu fungsi kerugian yang mengukur perbedaan antara distribusi probabilitas hasil prediksi model dan distribusi target sebenarnya pada setiap token. Loss ini bekerja dengan menghitung seberapa jauh probabilitas yang diberikan model terhadap token yang benar dibandingkan dengan label aslinya, sehingga semakin kecil nilainya semakin baik prediksi model. Cross-Entropy sangat sesuai untuk tugas language modeling dan summarization berbasis token karena output model berupa probabilitas atas seluruh kosakata.

Dilanjutkan dengan dilakukan nya training pada model dengan menggunakan 500 data sampel sebagai bahan dan menggunakan epoch sebanyak 3. Epoch adalah seberapa banyak model akan mengulangi 500 dataset itu dari awal sampai akhir. Detail Parameter yang dipakai dalam program adalah sebagai berikut :

Tabel 2. Parameter dan pengaruh

Parameter	Value dalam program	Pengaruh
Embedding Dimension (EMBED_DIM)	128	Menentukan kedalaman representasi kata; semakin besar semakin kaya informasi semantik, tetapi komputasi lebih berat.
Hidden Dimension BiLSTM (HIDDEN_DIM)	128	Menentukan kapasitas model menangkap konteks; semakin besar semakin mampu memahami dependensi panjang.
Jumlah Head Attention (nhead)	4	Menentukan jumlah pola relasi yang dapat dipelajari secara paralel dalam attention.
Jumlah Layer Transformer (num_layers)	1	Menentukan kedalaman abstraksi fitur; semakin banyak layer semakin kompleks representasi.
Sequence Length (SEQ_LEN)	256	Menentukan panjang maksimum input; semakin besar semakin banyak konteks yang bisa diproses.

Berdasarkan hasil pelatihan yang ditampilkan, proses training model BiLSTM Transformer dilakukan menggunakan 500 sampel data, yang seluruhnya berhasil dimuat dan diproses tanpa kendala. Setiap epoch terdiri dari 63 iterasi (batch), yang menunjukkan bahwa dataset dibagi secara konsisten untuk proses pembaruan bobot model. Waktu komputasi per epoch berada pada kisaran 3 menit 20 26 detik, dengan kecepatan pemrosesan sekitar 3,18 3,27 detik per iterasi, menandakan proses training berjalan stabil tanpa hambatan performa yang signifikan. Pada epoch pertama, model menghasilkan nilai average loss sebesar 8,1933, yang mencerminkan kondisi awal model yang masih dalam tahap penyesuaian terhadap pola data. Nilai loss yang relatif tinggi pada tahap ini menunjukkan bahwa bobot model belum optimal dalam merepresentasikan hubungan sekuensial dan kontekstual pada data teks. Hal ini merupakan kondisi yang wajar pada tahap awal pelatihan model deep learning. Memasuki epoch kedua, terjadi penurunan nilai average loss menjadi 7,0839, yang mengindikasikan bahwa model mulai belajar secara efektif dari data latih. Penurunan loss ini menunjukkan adanya peningkatan kemampuan model dalam menangkap dependensi lokal melalui BiLSTM serta konteks global melalui mekanisme self-attention pada Transformer. Proses pembaruan bobot berjalan konsisten, ditandai dengan stabilnya waktu iterasi dan tidak adanya lonjakan loss yang ekstrem. Pada epoch ketiga, nilai average loss kembali menurun menjadi 6,8608, menandakan konvergensi model yang semakin baik. Meskipun nilai loss pada iterasi terakhir epoch sedikit berfluktuasi, rata-rata loss tetap menunjukkan tren menurun secara keseluruhan. Hal ini mengindikasikan bahwa model telah mencapai kondisi pembelajaran yang relatif stabil dalam jumlah epoch yang terbatas, tanpa indikasi overfitting. Dengan demikian, hasil training ini menunjukkan bahwa arsitektur BiLSTM Transformer mampu mempelajari representasi teks secara efektif dan siap digunakan pada tahap evaluasi dan pengujian lebih lanjut. Model yang telah dilatih kemudian disimpan dalam bentuk file `bilstm_transformer_model.pt` untuk keperluan inferensi dan evaluasi lanjutan.

Selanjutnya adalah testing dengan memasukan link dari bbc secara langsung dahulu untuk melihat kinerja model.

ORIGINAL TEXT
The US will offer "more limited" support to allies, according to the Pentagon's new National Defense Strategy. In a significant shift to its security priorities, the US Department of Defense now considers security of the US homeland and Western Hemisphere - not China - as its primary concern. Previous versions of the strategy - published every four years - named the threat posed by China as the top defence priority. Relations with China will now be approached through "strength, not confrontation", the report says. The defence strategy reinforces recent calls from President Donald Trump, including for greater "burden-sharing" from allies in countering threats posed by Russia and North Korea. The new 34-page report follows last year's publication of the US National Security Strategy, which said Europe faced civilisational collapse and did not cast Russia as a threat to the US. At the time, Moscow said the document was "largely consistent" with its vision. By comparison, in 2018, the Pentagon described "revisionist powers", such as China and Russia, as the "central challenge" to US security. The new strategy calls on American allies to step up, saying partners have been "content" to let Washington subsidise their defence, although it denies the shift signals a US move towards "isolationism". "To the contrary, it means a focused and genuinely strategic approach to the threats our nation faces," it says. Washington has long neglected the "concrete interests" of Americans, the report says, adding the US does not want to conflate American interests "with those of the rest of the world - that a threat to a person halfway around the world is the same as to an American." Instead, it says allies, especially Europe, "will take the lead against threats that are less severe for us but more so for them". Russia, which launched a full-scale invasion of Ukraine nearly four years ago, is described as a "persistent but manageable threat to NATO's eastern members". Unlike in previous versions of the strategy, Taiwan, the self-governing island claimed by China, is not mentioned. However, the document does write that the US aims to "prevent anyone, including China, from being able to dominate us or our allies". Late last year, the US announced a vast arms sale to Taiwan worth \$11bn (£8.2bn), leading China to hold military drills around the island in response. The strategy also outlines a "more limited" role for US deterrence of North Korea. South Korea is "capable of taking primary responsibility" for the task, it adds. In the 12 months since Trump began his second term as president, the US has seized Venezuelan President Nicolás Maduro, carried out strikes against alleged drug boats in the eastern Pacific and Caribbean, and more recently, applied pressure on US allies to acquire Greenland. The strategy reiterates that the Pentagon "will guarantee US military and commercial access to key terrain, especially the Panama Canal, Gulf of America, and Greenland". The document says the Trump administration's approach will be "fundamentally different from the grandiose strategies of the past post-Cold War administrations". It adds: "Out with utopian idealism; in with hardnosed realism." At the World Economic Forum earlier this week, Trump claimed the US had "never gotten anything" from Nato and "we've never asked for anything". He further criticised the organisation, incorrectly claiming "the United States was paying for virtually 100% of Nato". Canadian Prime Minister Mark Carney said the old world order is "not coming back" and urged fellow middle powers - like South Korea, Canada and Australia - to come together. "Middle powers must act together because if we're not at the table, we're on the menu," Carney said at the Davos meeting. That came as French President Emmanuel Macron also warned of a "shift towards a world without rules".

Gambar 6. Awal Text

SUMMARY
Previous versions of the strategy - published every four years - named the threat posed by China as the top defence priority. Late last year, the US announced a vast arms sale to Taiwan worth \$11bn (£8.2bn), leading China to hold military drills around the island in response. South Korea is "capable of taking primary responsibility" for the task, it adds.

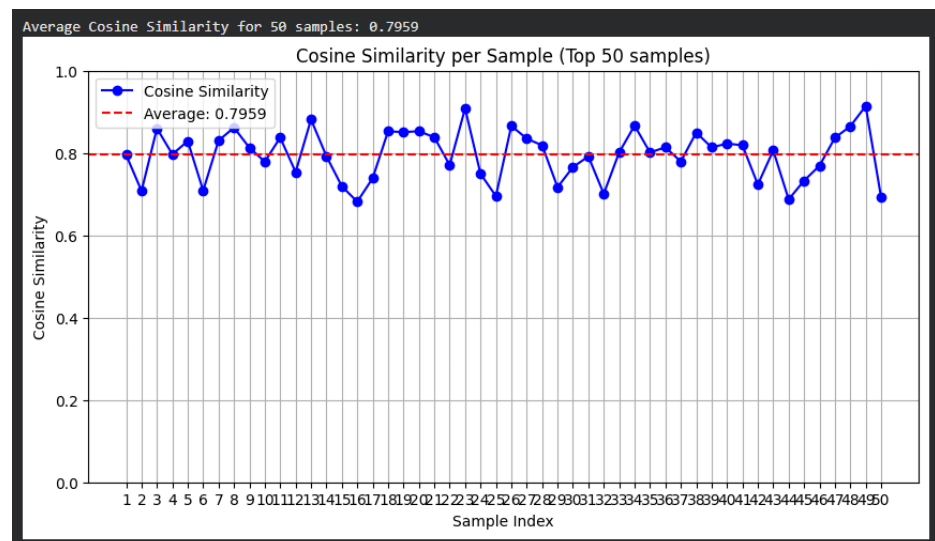
Gambar 7. Setelah Summarisasi

Link yang dipakai pada testing ini: <https://www.bbc.com/news/articles/cj9r8ezym3ro>

Secara kualitatif, hasil summarization ini memiliki kelebihan karena mampu mempertahankan beberapa informasi faktual penting, seperti isu prioritas terhadap China, penjualan senjata ke Taiwan, serta peran Korea Selatan dalam deterrence. Ringkasan juga tetap menggunakan kalimat asli sehingga tidak mengubah makna dasar dari informasi yang dipilih. Namun, kekurangannya adalah cakupan informasi yang sangat terbatas dan tidak merepresentasikan pergeseran besar strategi pertahanan AS secara menyeluruh. Koherensi antar kalimat kurang kuat karena ide yang dipilih terasa terpisah tanpa alur yang jelas. Selain itu, terdapat kegagalan dalam menampilkan angka secara rapi pada hasil summarizing, sehingga informasi numerik terlihat tidak terformat dengan baik. Keterbatasan cakupan informasi ini kemungkinan disebabkan oleh pembatasan panjang ringkasan yang diterapkan dalam proses summarizing.

4.2. Perbandingan Dataset lama dan Data baru

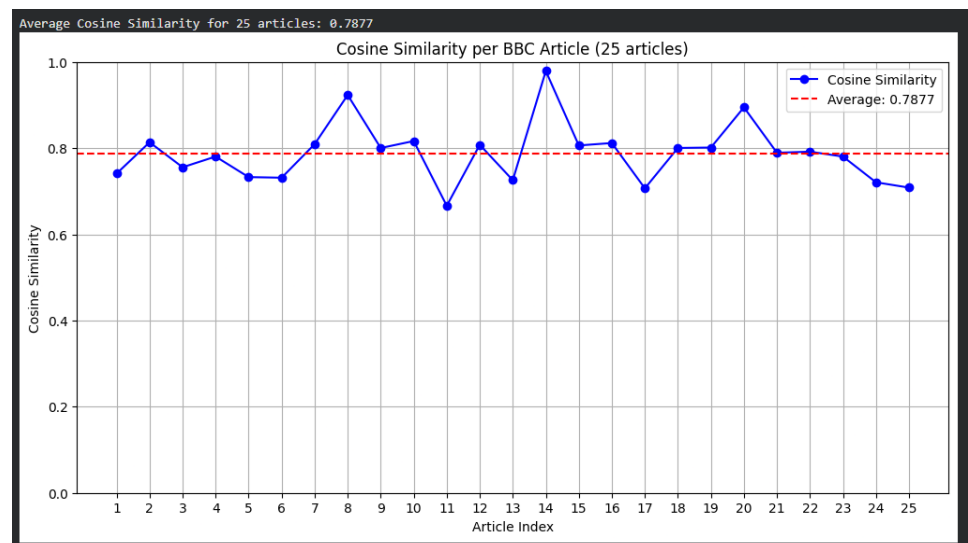
Pada tahap ini akan dilakukan evaluasi dengan cosine similarity untuk menunjukkan kinerja dari program dalam menjaga konteks kalimat. Untuk evaluasi ini akan dipakai 50 data untuk dites. Berikut adalah hasilnya



Gambar 8. Perubahan cosine similarity pada 50 sampel data dataset

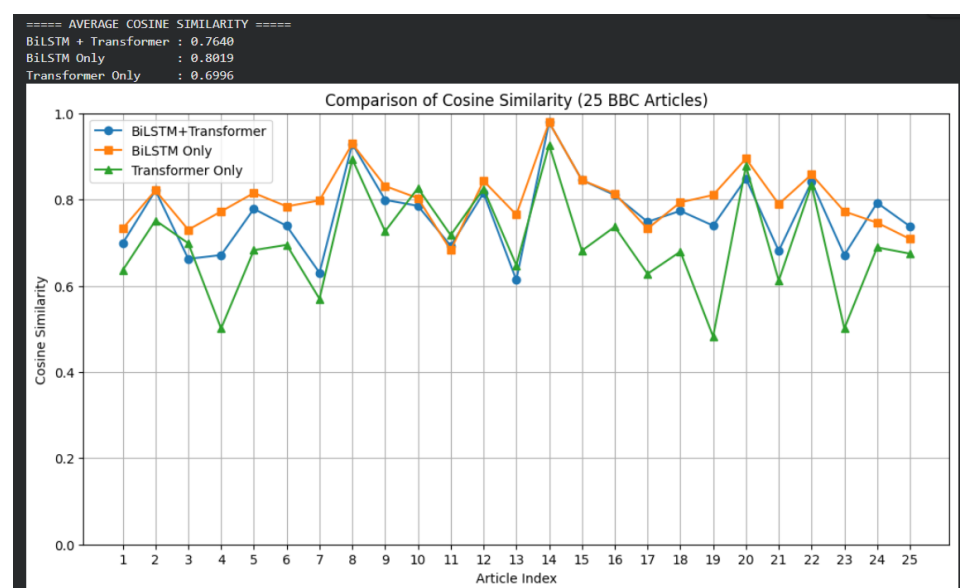
Gambar ini menampilkan cosine similarity untuk 50 sampel teratas dari dataset, yang menunjukkan seberapa mirip ringkasan yang dihasilkan model dengan teks aslinya. Setiap titik biru mewakili skor similarity dari satu sampel, sedangkan garis merah putus-putus menunjukkan nilai rata-rata similarity, yaitu sekitar 0,796. Terlihat bahwa sebagian besar sampel memiliki similarity di atas 0,7, menandakan model mampu menghasilkan ringkasan yang cukup representatif. Beberapa sampel mengalami penurunan skor di bawah rata-rata, yang menunjukkan adanya variasi dalam kualitas ringkasan antar sampel. Secara keseluruhan, grafik ini menggambarkan performa model yang konsisten dengan tingkat kesamaan yang cukup tinggi antara teks asli dan ringkasan yang dihasilkan.

Pada tahap akhir ini dilakukan perbandingan juga dengan cosine similarity dengan perbedaan kali ini menggunakan data terbaru bbc. berikut ini adalah hasil dari summarization 25 link bbc



Gambar 9. Perubahan cosine similarity pada data 25 link bbc

Hasil analisis menunjukkan bahwa rata-rata cosine similarity antara ringkasan dan teks asli dari 25 artikel BBC adalah sebesar 0,7877. Nilai ini mengindikasikan bahwa ringkasan yang dihasilkan model memiliki tingkat kesamaan semantik yang cukup tinggi dengan teks asli, sehingga model mampu menangkap inti informasi utama dengan baik. Variasi skor antar artikel terlihat signifikan, mulai dari sekitar 0,67 hingga hampir 1,0, yang menunjukkan bahwa beberapa artikel lebih mudah diringkas dibandingkan yang lain. Puncak tertinggi kemungkinan muncul pada artikel dengan struktur kalimat yang sederhana dan kosakata yang konsisten, sehingga model dapat memprediksi kata-kata penting lebih akurat. Sebaliknya, skor rendah kemungkinan disebabkan oleh artikel dengan bahasa kompleks, banyak istilah teknis, atau panjang artikel yang melebihi kapasitas model, sehingga ringkasan kehilangan beberapa konteks. Pola stabil di sekitar 0,78 menandakan konsistensi model dalam merangkum sebagian besar artikel, meskipun ada fluktuasi individual. Secara keseluruhan, hasil ini menunjukkan bahwa model BiLSTM + Transformer cukup efektif dalam membuat ringkasan yang relevan dan dapat diandalkan untuk teks berita BBC.



Gambar 10. Perbandingan antar ketiga model

Berdasarkan hasil pengujian pada 25 artikel BBC, model BiLSTM Only memperoleh rata-rata cosine similarity tertinggi sebesar 0.8019, diikuti oleh model Hybrid (BiLSTM + Transformer) sebesar 0.7640, dan Transformer Only sebesar 0.6996. Hasil ini menunjukkan bahwa pada kondisi data yang terbatas, BiLSTM mampu menangkap dependensi sekuensial dan pola lokal dalam teks secara lebih efektif dibandingkan mekanisme self-attention pada Transformer yang umumnya membutuhkan data lebih besar untuk bekerja optimal. Struktur berita yang relatif teratur dan memiliki pola naratif yang jelas juga mendukung keunggulan BiLSTM dalam memahami hubungan antar kata secara berurutan. Sementara itu, penambahan layer Transformer pada model Hybrid tidak memberikan peningkatan signifikan, kemungkinan karena kompleksitas tambahan tidak sebanding dengan jumlah data pelatihan yang tersedia. Dengan demikian, pada skenario dataset terbatas, model yang lebih sederhana seperti BiLSTM dapat memberikan performa yang lebih stabil dan konsisten dibandingkan arsitektur yang lebih kompleks.

6. Kesimpulan

Penelitian ini berhasil menunjukkan bahwa penggabungan arsitektur BiLSTM dan Transformer efektif dalam meningkatkan kualitas peringkasan teks otomatis pada berita BBC. Penggunaan lapisan BiLSTM terbukti mampu menangkap ketergantungan urutan kata secara lokal, sementara mekanisme self-attention pada Transformer memperkuat pemahaman konteks secara global. Implementasi pendekatan self-embedding memungkinkan model untuk bekerja secara optimal dalam kondisi tanpa label atau ground truth melalui metode unsupervised. Secara keseluruhan, sinergi kedua metode ini menghasilkan representasi teks yang informatif dan stabil, meskipun pada pengujian akhir model BiLSTM tunggal menunjukkan performa rata-rata tertinggi dibandingkan arsitektur lainnya.

Hasil evaluasi menggunakan metrik cosine similarity menunjukkan tingkat keberhasilan yang tinggi dengan nilai rata-rata keseluruhan berada pada kisaran 0,79, dengan BiLSTM mencapai 0,8019, model hibrida 0,7640, dan Transformer 0,6996. Temuan ini mengindikasikan bahwa pada kondisi dataset yang terbatas, pendekatan sekuensial BiLSTM lebih optimal dalam menangkap pola lokal teks berita yang terstruktur. Transformer yang mengandalkan mekanisme self-attention cenderung membutuhkan data yang lebih besar untuk mencapai performa maksimal, sehingga pada skenario ini belum menunjukkan keunggulan yang signifikan. Meskipun demikian, model hibrida tetap menunjukkan performa yang stabil dan konsisten baik saat diuji pada sampel dataset internal maupun pada artikel berita terbaru yang diambil langsung melalui tautan web. Mayoritas ringkasan yang dihasilkan tetap mampu menjaga inti informasi semantik dari teks asli meskipun terdapat fluktuasi pada artikel dengan struktur bahasa yang kompleks.

Ukuran dataset yang relatif terbatas, yaitu sekitar 500 sampel, berpotensi membatasi kemampuan model dalam menangkap variasi pola bahasa yang lebih luas. Dengan jumlah data yang tidak terlalu besar, model mungkin belum sepenuhnya memaksimalkan potensi arsitektur Transformer yang secara umum membutuhkan skala data lebih besar. Selain itu, ketergantungan utama pada cosine similarity sebagai mekanisme evaluasi dapat membuat penilaian terlalu berfokus pada kesamaan vektor permukaan dibanding kedalaman makna semantik yang lebih kompleks. Hal ini berpotensi mengabaikan nuansa konteks atau implikasi informasi yang tidak tercermin secara langsung dalam representasi numerik. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar serta menambahkan metrik evaluasi seperti ROUGE atau uji statistik untuk memperoleh gambaran performa yang lebih komprehensif dan meningkatkan kemampuan generalisasi model pada domain yang berbeda.

Kontribusi Penulis : "Konseptualisasi: Rafael Austin, Michael Jeconiah Yonathan, Andhika Dwi Rachmawanto, M Naufal Arriz; Metodologi: Michael Jeconiah Yonathan; Perangkat Lunak: Rafael Austin; Validasi: Michael Jeconiah Yonathan, Rafael Austin, dan Andhika Dwi Rachmawanto; Analisis Formal: Andhika Dwi Rachmawanto; Investigasi: Rafael Austin; Sumber Daya: Andhika Dwi Rachmawanto; Kurasi Data: Rafael Austin; Penulisan—penyusunan draf awal: Michael Jeconiah Yonathan; Penulisan—tinjauan dan penyuntingan: Michael Jeconiah Yonathan, Rafael Austin, dan Andhika Dwi Rachmawanto; Visualisasi: Rafael Austin; Supervisi: Andhika Dwi Rachmawanto; Administrasi proyek: Michael Jeconiah Yonathan; Pendanaan: Tidak ada"

Pendanaan : "Penelitian ini tidak menerima pendanaan eksternal, baik dari lembaga maupun individu. Penelitian ini murni merupakan hasil kerja independen dari tiap penulis."

Pernyataan Ketersediaan Data : "Data yang digunakan dalam penelitian ini adalah dataset publik yang tersedia secara bebas di huggingface. Dataset tersebut berjudul "Xsum" dan dapat diakses melalui tautan berikut: <https://huggingface.co/datasets/EdinburghNLP/xsum>. tidak ada data baru yang dibuat atau diarsipkan secara publik dalam penelitian ini."

Ucapan Terima Kasih : "Para penulis menyatakan bahwa dalam proses penulisan artikel ini, tidak ada penggunaan alat kecerdasan buatan (AI) untuk membuat konten teks, gambar, atau data yang disajikan."

Konflik Kepentingan : "Para penulis menyatakan tidak ada konflik kepentingan."

Daftar Pustaka

- [1] Halimah, Surya Agustian, and Siti Ramadhani, 'Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrak', *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 371 381, Dec. 2022, doi: 10.37859/coscitech.v3i3.4300.
- [2] V. Gupta and G. S. Lehal, 'A Survey of Text Summarization Extractive techniques', in *Journal of Emerging Technologies in Web Intelligence*, Aug. 2010, pp. 258 268. doi: 10.4304/jetwi.2.3.258-268.
- [3] R. Patil, S. Boit, V. Gudivada, and J. Nandigam, 'A Survey of Text Representation and Embedding Techniques in NLP', doi: 10.1109/ACCESS.2022.0092316.
- [4] Y. Asri, D. Kuswardani, A. A. Sari, and A. R. Ansari, 'Word embedding for contextual similarity using cosine similarity', *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 2, p. 1170, May 2025, doi: 10.11591/ijeecs.v38.i2.pp1170-1180.
- [5] Halimah et al, "Peringkasan teks otomatis (automated text summarization) pada artikel berbahasa indonesia menggunakan algoritma lexrak", vol 3, no 3, 2022, doi: <https://doi.org/10.37859/coscitech.v3i3.4300>
- [6] C. Raffel *et al.*, 'Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer', 2020. [Online]. Available: <http://jmlr.org/papers/v21/20-074.html>.
- [7] A. See, P. J. Liu, and C. D. Manning, 'Get to the point: Summarization with pointer-generator networks', in *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, Association for Computational Linguistics (ACL), 2017, pp. 1073 1083. doi: 10.18653/v1/P17-1099.
- [8] V. Tundjungsari, Dasar Machine Learning, Edisi Revisi. Yogyakarta, Indonesia: Deepublish, 2024.
- [9] N. Q. Habban and M. H. Abdulameer, 'A Hybrid Model for Long Document of Academic Research Papers Summarization Using BiLSTM-based Sentence Classification and Hierarchical Attention Mechanism', *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 7, pp. 635 654, 2025, doi: 10.22266/ijies2025.0831.41.
- [10] A. Vaswani et al., "Attention Is All You Need," in Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS), 2020, pp. 6000 6010, arXiv:2003.07000v1.
- [11] Z. Huang, P. Xu, D. Liang, A. Mishra, and B. Xiang, 'TRANS-BLSTM: Transformer with Bidirectional LSTM for Language Understanding', Mar. 2020, [Online]. Available: <http://arxiv.org/abs/2003.07000>
- [12] B. Rahmadhani, P. Purwono, and Safar Dwi Kurniawan, 'Understanding Transformers: A Comprehensive Review', *Journal of Advanced Health Informatics Research*, vol. 2, no. 2, pp. 85 94, Aug. 2024, doi: 10.59247/jahir.v2i2.292.
- [13] A. Vaswani *et al.*, 'Attention Is All You Need', 2023.
- [14] A. Chechkin et al, "A Hybrid Neural Network Transformer for Detecting and Classifying Destructive Content in Digital Space", MDPI, 2025, DOI : <https://doi.org/10.3390/a18120735>
- [15] G. Tamrakar, 'A Comparative Analysis of Transformer and LSTM Architectures for Text Summarization: A Case Study on News and Scientific Article Corpora', *SECITS Journal of Scalable Distributed Computing and Pipeline Automation*, vol. 1, no. 1, pp. 24 31, doi: 10.17051/SJSDCPA/01.01.04.
- [16] F. Liu, N. Zhao, and G. Zhu, 'Cognitive difference text classification in online knowledge collaboration based on SA-BiLSTM hybrid model', *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-06914-w.
- [17] A. Pannadhika Putra, D. Purnami Singgih Putri, and Aak. Cahyawan Wiranatha, 'Scientific Paper Recommendation System: Application of Sentence Transformers and Cosine Similarity Using arXiv Data', 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [18] Y. M. Wazery, M. E. Saleh, and A. A. Ali, 'An optimized hybrid deep learning model based on word embeddings and statistical features for extractive summarization', *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 7, Jul. 2023, doi: 10.1016/j.jksuci.2023.101614.
- [19] A. Chechkin, E. Pleshakova, and S. Gataullin, 'A Hybrid Neural Network Transformer for Detecting and Classifying Destructive Content in Digital Space', *Algorithms*, vol. 18, no. 12, p. 735, Nov. 2025, doi: 10.3390/a18120735.
- [20] G Tamrakar, "A Comparative Analysis of Transformer and LSTM Architectures for Text Summarization : A Case Study on News and Scientific Article Corpora", SECITS, Vol 1 No.1, 2024, DOI : <https://doi.org/10.17051/SJSDCPA/01.01.04>