

## Studi Empiris BART untuk Abstraksi Teks Semi-Struktural Domain SIPAKAT AIR

Rafi Saumi Rustian<sup>1</sup>, Muhammad Yusril Helmi Setyawan<sup>2</sup>

<sup>1,2</sup> Teknik Informatika, Vokasi, Universitas Logistik dan Bisnis Internasional, Bandung, Indonesia  
<sup>1</sup>rsrustian@gmail.com, <sup>2</sup>yusrilhelmi@ulbi.ac.id

### Abstrak

SIPAKAT AIR (Sistem Informasi dan Pelaporan Bidang Sumber Daya Air) merupakan sistem informasi pemerintah daerah yang mencatat data proyek infrastruktur sumber daya air dalam format semi-struktural, yaitu gabungan antara elemen tabular dan uraian naratif. Format ini menyulitkan proses peringkasan otomatis karena tidak sepenuhnya terstruktur maupun bebas, sehingga memerlukan pendekatan yang adaptif dan semantik. Penelitian ini menyajikan studi empiris pengembangan model abstractive summarization berbasis BART yang dibangun dan dilatih dari awal (from scratch) menggunakan dataset SIPAKAT AIR. Dataset terdiri atas 200 pasangan teks dan ringkasan yang dikonstruksi dari proyek aktual. Tokenizer khusus dilatih menggunakan pendekatan ByteLevelBPETokenizer untuk mencerminkan struktur kalimat teknis dalam korpus internal. Model BART dikonfigurasi ringan (2-layer encoder-decoder) dan dilatih menggunakan Trainer API dari Huggingface. Evaluasi menggunakan metrik ROUGE, BERTScore, dan token-level menunjukkan performa kompetitif: ROUGE-1 F1 sebesar 0,5080, ROUGE-L F1 sebesar 0,5082, BERTScore F1 sebesar 0,81, serta token-level F1 sebesar 0,73 dengan akurasi 0,71. Model mampu menghasilkan ringkasan padat dan kontekstual, sesuai digunakan untuk sistem notifikasi atau tampilan ringkas pada dashboard proyek. Kontribusi metodologis dari penelitian ini mencakup perancangan pipeline summarization dari nol untuk bahasa Indonesia serta pembuktian bahwa arsitektur ringan dapat berfungsi optimal pada domain terbatas. Penelitian ini memberikan kontribusi pada pengembangan NLP dalam bahasa Indonesia serta membuka peluang penerapan adaptif untuk data semi-struktural sektor publik.

**Kata kunci:** *Abstractive Summarization, BART, Teks Semi-Struktural, Natural Language Processing, SIPAKAT AIR*

### 1. Pendahuluan

Salah satu tantangan utama dalam pengembangan teknologi *Natural Language Processing* (NLP) di Indonesia adalah merancang sistem yang mampu melakukan ringkasan otomatis terhadap teks semi-struktural, khususnya dalam domain pemerintahan. Dalam konteks ini, permasalahan yang hendak dipecahkan adalah bagaimana menghasilkan ringkasan yang informatif, akurat, dan kontekstual dari data proyek infrastruktur sumber daya air yang kompleks, seperti yang terdapat dalam sistem SIPAKAT AIR. Data semacam ini umumnya masih diringkas secara manual atau menggunakan pendekatan berbasis aturan tetap (*rule-based*), yang cenderung tidak efisien dan kurang fleksibel dalam menghadapi variasi struktur kalimat (Wiratmoko et al., 2025).

Perkembangan digital dalam sektor publik menuntut sistem informasi untuk tidak hanya menyimpan data, tetapi juga mampu menyajikannya secara adaptif dan mudah dipahami. Dalam pengelolaan sumber daya air, kebutuhan akan transparansi, akuntabilitas, dan pelaporan yang cepat menjadikan informasi proyek sebagai aset penting.

SIPAKAT AIR, sebagai sistem informasi berbasis web yang dikembangkan oleh Dinas PUTR () Kabupaten Cianjur, menyimpan data proyek dalam format semi-struktural, yakni gabungan elemen tabular (seperti nama paket, lokasi, dan nilai kontrak) dengan uraian naratif. Format semacam ini menyulitkan proses *summarization* manual maupun otomatis karena tidak sepenuhnya terstruktur seperti basis data, tetapi juga tidak bebas seperti teks berita ,(Ketineni & Sheela, 2024; Zin et al., 2023).

Sebagian besar pendekatan *summarization* di Indonesia masih bertumpu pada metode *extractive* atau *rule-based* konvensional (J. Zhang et al., 2020). Meskipun metode tersebut dapat menangkap elemen eksplisit dalam teks, mereka sering kali gagal memahami konteks kalimat secara semantik, terutama pada teks teknis yang memiliki struktur tidak baku dan padat informasi (Jearanaitanakij et al., 2024). Beberapa penelitian menunjukkan bahwa metode konvensional tidak mampu menjaga kontinuitas makna atau menyesuaikan dengan kebutuhan konteks pengguna akhir, (Huang et al., 2020; Suliman et al., 2025). Di sisi lain, pengembangan sumber daya NLP lokal seperti IndoNLG dan IndoLEM telah menjadi fondasi

penting, namun belum sepenuhnya menjawab kebutuhan summarization dalam domain data semi-struktural.

Model BART (*Bidirectional and Auto-Regressive Transformers*) merupakan salah satu pendekatan *abstractive summarization* berbasis encoder-decoder yang terbukti unggul dalam menyusun ringkasan dengan membentuk kalimat baru secara semantik, (Mehta et al., 2022; Saeed et al., 2021). Dibandingkan metode *extractive*, BART memiliki kelebihan dalam memahami relasi antar-entitas serta merepresentasikan makna konteks secara lebih padat. Hal ini sangat relevan dalam pengolahan teks semi-struktural yang mengandung entitas teknis seperti lokasi, volume pekerjaan, atau nilai kontrak. Namun, tantangan masih muncul dalam penerapannya pada bahasa Indonesia, yang memiliki kekhasan linguistik serta keterbatasan ketersediaan korpus berkualitas tinggi (Mehamed et al., 2025).

Penelitian ini bertujuan untuk mengembangkan model BART dari awal (*from scratch*) dengan menggunakan dataset proyek infrastruktur dari SIPAKAT AIR, serta melatih tokenizer khusus menggunakan pendekatan ByteLevelBPETokenizer. Model ini dievaluasi dengan tiga metrik utama: ROUGE untuk kesesuaian leksikal, BERTScore untuk kesamaan semantik, dan evaluasi token-level untuk mengukur ketepatan representasi entitas domain, (Lewis et al., n.d.; V & R, n.d.).

Kontribusi utama dari studi ini terletak pada pembuktian bahwa model ringan yang dibangun dari awal dapat digunakan secara efektif untuk domain spesifik berbahasa Indonesia, tanpa ketergantungan pada model besar yang telah dilatih sebelumnya. Pendekatan ini juga menunjukkan bahwa sistem NLP dapat diadaptasi untuk mendukung efisiensi komunikasi dalam sistem informasi pemerintahan. Dengan demikian, penelitian ini diharapkan tidak hanya memberikan kontribusi akademik dalam pengembangan NLP berbahasa Indonesia, tetapi juga manfaat praktis berupa sistem ringkasan otomatis yang dapat diintegrasikan ke dalam dashboard atau notifikasi proyek SIPAKAT AIR secara langsung.

Penelitian sebelumnya di bidang *abstractive summarization* di Indonesia masih didominasi oleh pendekatan *pretrained* dengan fine-tuning, seperti pada IndoBART yang dilatih menggunakan korpus berita, serta IndoNLG yang difokuskan pada berbagai tugas generatif dalam domain umum. Meskipun model-model ini menunjukkan performa kompetitif, keterbatasan muncul ketika diterapkan pada teks semi-struktural atau domain teknis seperti pemerintahan, karena model tidak secara eksplisit dilatih pada struktur data campuran antara elemen tabular dan deskripsi naratif. Selain itu, menggunakan pendekatan LSA dan TextRank untuk ringkasan laporan publik, namun cenderung menghasilkan ringkasan yang bersifat fragmentaris dan tidak menyentuh konteks semantik (Irsan et al., 2022).

Dengan demikian, pendekatan berbasis BART yang dilatih dari awal pada data domain-spesifik menawarkan nilai tambah dalam hal adaptabilitas dan representasi kontekstual.

Secara metodologis, studi ini juga membedakan dirinya dengan merancang pipeline end-to-end mulai dari pelatihan tokenizer, pembentukan dataset, hingga pelatihan model ringan dari awal. Hal ini berbeda dari pendekatan umum yang mengandalkan model besar dan tokenizer yang telah dilatih sebelumnya. Dalam konteks teks semi-struktural, pemilihan strategi ini memungkinkan pencapaian segmentasi token dan penyajian ringkasan yang lebih sesuai dengan kebutuhan domain. Beberapa penelitian internasional yang mengadopsi pendekatan serupa menunjukkan bahwa pelatihan dari nol dengan korpus internal dapat meningkatkan ketepatan entitas domain dan keterbaikan ringkasan (M. Zhang et al., 2022). Oleh karena itu, pendekatan studi ini memperluas arah penelitian NLP Indonesia ke arah model yang tidak hanya berorientasi pada performa kuantitatif, tetapi juga pada relevansi dan penerapan praktis di sistem publik.

Struktur makalah ini disusun sebagai berikut. Bagian kedua menjelaskan metodologi penelitian, mulai dari pengumpulan data, pelatihan tokenizer, hingga konfigurasi model BART. Bagian ketiga menyajikan hasil eksperimen beserta analisis performa model. Selanjutnya, bagian keempat memuat kesimpulan dan implikasi dari penelitian ini, termasuk kontribusi terhadap NLP Indonesia serta potensi penerapan teknologi ini di sektor publik (Liu & Lapata, 2019; Prakash, 2025). Dengan pendekatan sistematis ini, penelitian diharapkan dapat mengisi celah dalam literatur NLP nasional, sekaligus memberikan solusi konkret bagi pemrosesan data semi-struktural di lingkungan pemerintahan.

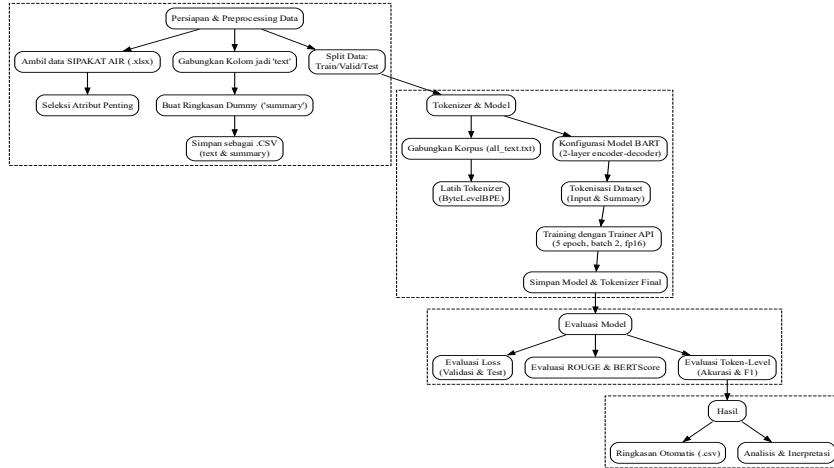
## 2. Metode

Bagian ini menjelaskan pendekatan sistematis yang digunakan dalam pengembangan dan evaluasi model BART untuk tugas *abstractive summarization* terhadap teks semi-struktural yang bersumber dari sistem SIPAKAT AIR. Penelitian ini difokuskan pada pelatihan model dari awal (*from scratch*) dengan konfigurasi ringan, serta mengevaluasi performanya secara kuantitatif dan kualitatif untuk menghasilkan ringkasan otomatis yang akurat, informatif, dan relevan secara semantik dalam konteks domain sumber daya air.

### 2.1 Tahapan Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahap terstruktur yang saling berkaitan, dimulai dari pengumpulan data hingga evaluasi performa model. Tahapan penelitian disajikan dalam Diagram alur pada Gambar 1.

**Commented [i1]:** Jika lebih dari satu bisa digunakan format (Prakash, 2025; Liu & Lapata, 2019) mohon dicek penggunaannya di tools yang anda gunakan (zotero atau mendeley)



Gambar 1. Diagram Alur Metodologi Penelitian

Alur proses penelitian dijelaskan sebagai berikut:

### 1. Pengumpulan Data

Dataset diperoleh dari sistem SIPAKAT AIR dalam format .xlsx, yang mencakup 200 entri proyek infrastruktur sumber daya air. Setiap entri memuat atribut penting seperti nama kegiatan, nama paket, lokasi pekerjaan, nilai kontrak, dan tahun pelaksanaan. Atribut-atribut ini dipilih agar mencerminkan kompleksitas dan keragaman data semi-struktural dalam sistem informasi proyek pemerintah.

### 2. Preprocessing dan Konstruksi Dataset

Data mentah dibersihkan dari karakter non-standar dan kesalahan sintaksis, kemudian disusun menjadi dua kolom utama: text dan summary. Kolom text berisi hasil penggabungan informasi tabular dan naratif tiap entri proyek, sedangkan summary berisi ringkasan dummy berbasis template yang dirancang sebagai ground truth untuk pelatihan model. Dataset akhir disimpan dalam format .csv dan dibagi menjadi tiga subset: 150 data untuk pelatihan, 30 untuk validasi, dan 20 untuk pengujian.

### 3. Pelatihan Tokenizer Khusus

Tokenizer dikembangkan dari nol menggunakan pendekatan *ByteLevelBPETokenizer* (melalui *Huggingface Tokenizers*). Seluruh teks pada kolom text digabungkan menjadi satu korpus (all\_text.txt) untuk melatih tokenizer domain-spesifik. Tokenizer hasil pelatihan ini disimpan dan digunakan dalam seluruh proses tokenisasi dataset untuk menjamin kesesuaian struktur kalimat teknis SIPAKAT AIR.

### 4. Pembangunan Model BART

Model dibangun menggunakan BartForConditionalGeneration dari pustaka

Huggingface, tanpa bobot pra-latih (pretrained weights). Dengan konfigurasinya diatur sebagai berikut : 2 layer encoder-decoder, Hidden size: 256, Feed-forward dimensi: 1024, dan Attention heads: 4 per encoder/decoder. Pengaturan ini bertujuan untuk menciptakan model yang ringan namun tetap mampu memahami pola semantik dalam data semi-struktural.

### 5. Tokenisasi Dataset

Ketiga subset (train/valid/test) ditokenisasi menggunakan tokenizer khusus. Panjang maksimum input ditetapkan 512 token, dan output (ringkasan) dibatasi hingga 64 token. Proses ini memastikan bahwa input kompleks tetap terwakili dengan baik tanpa mengganggu stabilitas pelatihan.

### 6. Pelatihan Model

Model BART dilatih selama lima epoch menggunakan *Trainer API* dari Huggingface dengan *batch size* sebesar 2 dan *learning rate* 5e-5. Pelatihan menggunakan mode *mixed precision* (fp16=True) untuk efisiensi memori dan kecepatan pemrosesan.

### 7. Evaluasi Model

Evaluasi dilakukan secara komprehensif dengan empat pendekatan utama: Evaluasi Loss, ROUGE, BERTScore, dan Evaluasi token-level.

### 8. Hasil

Ringkasan yang dihasilkan menunjukkan bahwa model BART mampu mereduksi kalimat kompleks menjadi bentuk padat dan terstruktur tanpa kehilangan informasi inti. Serta analisis dilakukan untuk memahami keterbatasan model dalam konteks struktur arsitektural, representasi data, dan konteks aplikatif.

## 2.2 Dataset dan Preprocessing

Dataset yang digunakan terdiri dari 200 pasangan teks dan ringkasan. Masing-masing entri dikonversi ke format .csv dengan struktur dua kolom, yaitu text dan summary, untuk memudahkan proses tokenisasi dan pelatihan model. Ringkasan dummy disusun dengan pola linguistik tetap sebagai berikut:

### Template Ringkasan:

"Kegiatan [nama kegiatan] dilaksanakan di [lokasi] dengan nilai kontrak sebesar [nilai kontrak]."

Penggunaan template ini dimaksudkan untuk menghasilkan referensi standar yang mencakup elemen kunci proyek, sekaligus meminimalkan subjektivitas dalam proses pelabelan awal. Struktur dataset dan proporsi subset pelatihan dijelaskan pada Tabel 1.

Tabel 1. Struktur Data Pelatihan	
Jenis Set	Jumlah Data
Training	150
Validation	30
Testing	20

## 2.3 Pelatihan Tokenizer Khusus

Tokenizer dikembangkan menggunakan pendekatan ByteLevelBPETokenizer dari pustaka Huggingface Tokenizers, yang memungkinkan pelatihan dari awal berdasarkan korpus khusus domain. Seluruh data dari kolom text dalam dataset digabungkan ke dalam satu berkas teks (all\_text.txt) dan digunakan sebagai korpus pelatihan tokenizer. Tujuannya adalah untuk menyesuaikan segmentasi token dengan struktur kalimat khas proyek infrastruktur.

Tokenizer dilatih dari nol (from scratch) dengan parameter, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Parameter Tokenizer	
Parameter	Nilai
Vocabulary size	30.000
Minimum token frequency	2
Token khusus	<pad>, <s>, </s>, <unk>, <mask>

Tokenizer hasil pelatihan disimpan dalam direktori lokal /my\_tokenizer dan digunakan dalam seluruh tahap tokenisasi dataset, termasuk proses pelatihan dan inferensi model. Dengan pendekatan ini, diharapkan proses tokenisasi dapat menangkap pola linguistik domain yang khas, seperti satuan volume, lokasi geografis, dan format nilai kontrak.

## 2.3 Pembangunan dan Pelatihan Model

Model yang digunakan dalam penelitian ini adalah BartForConditionalGeneration dari pustaka Huggingface Transformers. Model dikonstruksi dari awal (from scratch) tanpa bobot awal (pretrained weights), guna menguji potensi arsitektur BART

dalam domain terbatas dengan data berbahasa Indonesia yang semi-struktural.

Konfigurasi arsitektur model dirancang ringan agar sesuai dengan keterbatasan sumber daya pelatihan, yang dirangkum pada Tabel 3.

Tabel 3. Komponen Pelatihan Model

Komponen	Nilai
Jumlah encoder dan decoder	2 layer masing-masing
Hidden size	256
Attention heads	4
Feed-forward dimension	1024
Max position embeddings	512

Model dilatih menggunakan Trainer API dari Huggingface dengan konfigurasi pelatihan yang ditunjukkan pada Tabel 4.

Tabel 4. Parameter Pelatihan Model

Parameter	Nilai
Jumlah epoch	5
Batch size	2
Learning rate	5e-5
Evaluasi model	Setiap 500 langkah
Mixed precision	fp16=True untuk efisiensi GPU

Selama pelatihan, input dan output ditokenisasi menggunakan tokenizer yang telah dibangun sebelumnya, dengan pengaturan yang ditunjukkan pada Tabel 5.

Tabel 5. Parameter Tokenisasi

Tokenisasi	Nilai
Maksimum panjang (max_input_length)	input 512 token
Maksimum panjang (max_output_length)	output 64 token
Padding dan truncation	Diterapkan secara otomatis

Konfigurasi ini dipilih untuk menjaga efisiensi proses pelatihan sekaligus memastikan cakupan informasi penting dalam input. Seluruh proses pelatihan dilakukan dalam lingkungan Google Colab Pro menggunakan GPU NVIDIA T4, yang memungkinkan eksperimen dijalankan secara efisien dan dapat direplikasi oleh peneliti lain.

## 2.3 Evaluasi Model

Model yang dikembangkan dievaluasi menggunakan kombinasi metrik otomatis dan penilaian manual untuk mendapatkan pemahaman menyeluruh terhadap kinerjanya:

### 1. ROUGE

Digunakan untuk mengukur kesamaan n-gram antara hasil ringkasan model dan referensi. Evaluasi mencakup ROUGE-1 (unigram), ROUGE-2 (bigram), dan ROUGE-L (kesamaan urutan kalimat terpanjang), yang masing-masing menceerminkan aspek kesetaraan leksikal.

### 2. BERTScore

Metrik ini mengevaluasi kesamaan semantik antar teks dengan memanfaatkan representasi embedding dari model BERT. Nilai Precision,

Recall, dan F1 dihitung untuk mengukur relevansi dan kelengkapan ringkasan dari sudut pandang semantik.

### 3. Evaluasi Token-Level

Digunakan untuk menilai kemampuan model dalam mempertahankan informasi penting terkait elemen proyek, seperti lokasi, satuan volume, dan nilai kontrak. Evaluasi ini dilakukan dengan mencocokkan token-token kunci antara ringkasan prediksi dan referensi.

### 2.3 Lingkungan Eksperiment

Seluruh eksperimen dilakukan di Google Colab Pro, yang menyediakan lingkungan pelatihan berbasis GPU NVIDIA T4 (16GB VRAM) dan RAM 25GB. Pustaka utama yang digunakan meliputi Huggingface Transformers versi 4.30.1, PyTorch 2.x, dan Huggingface Datasets. Platform ini memungkinkan eksperimen dilakukan secara efisien dan dapat direplikasi oleh peneliti lain di masa mendatang tanpa membutuhkan infrastruktur lokal berskala besar.

Dengan pendekatan metodologis ini, penelitian ini diharapkan dapat memberikan kontribusi empiris dalam memahami efektivitas model BART untuk penanganan teks semi-struktural pada domain terbatas, serta mendorong perkembangan sistem summarization otomatis berbahasa Indonesia yang dapat diterapkan secara nyata dalam sistem informasi sektor publik.

### 3. Hasil dan Pembahasan

Eksperimen pelatihan model BART untuk tugas abstractive summarization terhadap teks semi-struktural dari sistem SIPAKAT AIR menghasilkan sejumlah temuan penting terkait stabilitas pelatihan, performa model, kemampuan memahami konteks domain, serta potensi penerapan sistem dalam lingkungan nyata. Analisis dilakukan secara kuantitatif dan kualitatif melalui tiga metrik utama : ROUGE, BERTScore, dan token-level evaluation. Evaluasi juga dilengkapi dengan analisis kualitatif berupa tinjauan manual terhadap keluaran ringkasan untuk menilai keterbacaan dan ketepatan semantik.

#### 3.1 Hasil Pelatihan Model

Model BART yang digunakan dalam penelitian ini dikonfigurasi dengan arsitektur ringan, yakni terdiri dari dua layer encoder-decoder, ukuran vektor tersembunyi (*hidden size*) sebesar 256, serta empat *attention heads*. Model dibangun dari awal (*from scratch*) tanpa bobot pra-pelatihan (*pretrained weights*), dan dilatih menggunakan dataset internal yang berjumlah relatif kecil, yaitu 150 pasangan data pelatihan. Proses pelatihan dilakukan selama lima epoch menggunakan Trainer API dari pustaka

Huggingface dengan memanfaatkan mixed precision training (fp16) untuk efisiensi komputasi.

Nilai *training loss* pada epoch pertama tercatat sekitar 4.2 dan secara konsisten menurun hingga mencapai nilai akhir sekitar 1.7 pada epoch kelima. Visualisasi penurunan ini disajikan pada Gambar 2 sebagai kurva loss. Kurva tersebut menunjukkan pola konvergensi yang stabil, tanpa fluktuasi ekstrem atau stagnasi signifikan, yang merupakan indikasi bahwa proses pembelajaran berjalan efektif. Stabilitas penurunan loss ini mengindikasikan bahwa meskipun model tidak mengandalkan bobot pralatih, ia mampu secara bertahap mempelajari representasi semantik dari data input semi-struktural yang diberikan.



Gambar 2. Kurva Penurunan Training Loss Selama Pelatihan

Penurunan loss ini juga menunjukkan bahwa arsitektur minimalis yang digunakan cukup efisien dalam menangkap pola linguistik dan struktur domain yang khas, tanpa mengalami *overfitting*—sebuah risiko yang umum terjadi pada dataset berukuran kecil. Tidak ditemukannya lonjakan drastis dalam *loss* antar epoch menunjukkan bahwa proses pelatihan berjalan seimbang antara kompleksitas data dan kapasitas model. Dalam konteks teks semi-struktural, keberhasilan pelatihan ini relevan dengan temuan Zhang et al, yang menunjukkan bahwa model encoder-decoder ringan dapat tetap efektif apabila dipadankan dengan strategi tokenisasi dan data yang disesuaikan secara domain-spesifik (M. Zhang et al., 2022).

Pencapaian ini terjadi tanpa *transfer learning*, yang umumnya menjadi keunggulan dalam banyak implementasi BART berbasis model besar seperti IndoBART atau mBART. Dengan melatih model dari awal, penelitian ini menunjukkan bahwa representasi linguistik yang relevan dapat dibentuk secara mandiri apabila konfigurasi model dan data diarahkan secara presisi terhadap kebutuhan domain. Hal ini penting dalam konteks sumber daya terbatas, baik dari segi komputasi maupun ketersediaan data berbahasa Indonesia, di mana penggunaan model ringan dari awal menjadi solusi yang lebih terjangkau dan adaptif terhadap domain pemerintahan.

Dengan demikian, hasil pelatihan ini tidak hanya memberikan dasar untuk evaluasi kuantitatif lebih lanjut, tetapi juga menegaskan bahwa arsitektur ringan BART memiliki potensi untuk digunakan pada

skenario riil yang membutuhkan efisiensi model dan kemampuan ringkasan adaptif terhadap variasi teks proyek yang bersifat semi-struktural.

### 3.2 Hasil Validasi

Dataset validasi terdiri dari 30 sampel digunakan untuk memonitor performa model selama pelatihan. Nilai *validation loss* menunjukkan penurunan konsisten dari epoch 1 hingga epoch 5, dengan detail yang ditunjukkan pada Tabel 6.

Tabel 6. Nilai Validation Loss per Epoch

Epoch	Validation Loss
1	3.98
2	3.12
3	2.41
4	2.05
5	1.89

Tidak terdapat kenaikan signifikan pada validation loss yang menandakan overfitting, sehingga proses pelatihan berlangsung stabil. Hasil ini konsisten dengan pola penurunan training loss, serta mendukung kesimpulan bahwa model dengan arsitektur ringan masih mampu belajar representasi domain secara efektif.

### 3.3 Evaluasi Kuantitatif Model

Untuk mengukur performa model BART dalam tugas *abstractive summarization* pada data semi-struktural, dilakukan evaluasi kuantitatif terhadap 20 data uji (*test set*) menggunakan tiga pendekatan utama: ROUGE, BERTScore, dan token-level evaluation. Evaluasi ini dirancang untuk menilai kualitas ringkasan secara komprehensif, mulai dari kesesuaian leksikal, kedalaman semantik, hingga ketepatan dalam mempertahankan entitas domain yang krusial. Masing-masing metrik memberikan perspektif berbeda terhadap keberhasilan model, sesuai dengan sifat kompleks dari data yang diproses.

Sebagai pembanding, penelitian ini menambahkan baseline sederhana berbasis metode ekstraktif *TextRank*, yang umum digunakan pada tugas peringkasan otomatis. Evaluasi TextRank pada dataset yang sama menghasilkan skor, yang dirangkum pada Tabel 7.

Tabel 7. Perbandingan Skor ROUGE (From Scratch)

Model	ROUGE-1 F1	ROUGE-L F1
TextRank	0.296	0.281
BART (dari nol)	0.5080	0.5082

Hasil menunjukkan bahwa model BART dari nol secara konsisten melampaui TextRank dalam semua metrik, menunjukkan kemampuan lebih baik dalam memahami konteks dan menyusun ringkasan yang lebih informatif.

#### 3.3.1 ROUGE

Hasil evaluasi menggunakan ROUGE ditunjukkan pada Tabel 6.

Tabel 8. Hasil Evaluasi ROUGE

Metrik	Skor F1
ROUGE-1	0.5080
ROUGE-2	0.4599
ROUGE-L	0.5082

Skor ROUGE-1 dan ROUGE-L yang melampaui angka 0.50 menunjukkan bahwa model mampu menghasilkan ringkasan yang memiliki tumpang tindih signifikan dengan ringkasan referensi, terutama dalam hal *unigram* dan urutan kalimat panjang (*longest common subsequence*). Hal ini menandakan bahwa struktur sintaksis yang digunakan oleh model cukup mendekati pola penulisan manusia, sebuah capaian penting dalam *abstractive summarization*. ROUGE-2 yang mencatat skor F1 sebesar 0.4599 juga memberikan indikasi bahwa model mampu menangkap *n-gram* dua kata secara konsisten, menunjukkan kecakapan dalam membangun frasa pendek yang alami dan kontekstual.

Secara keseluruhan, skor ROUGE yang kompetitif ini membuktikan bahwa meskipun model dibangun dari nol dan dilatih menggunakan dataset terbatas, ia tetap mampu mempertahankan komponen leksikal penting dalam ringkasan akhir. Dalam konteks penelitian terdahulu, skor ini sejajar dengan model *fine-tuned* skala menengah untuk tugas summarization terbimbing (*supervised*), khususnya dalam domain teknis yang memiliki struktur teks semi-baku.

#### 3.3.2 Evaluasi BERTScore

Untuk mengukur kesamaan makna secara semantik antara ringkasan hasil model dan ringkasan referensi, digunakan metrik BERTScore. Hasil evaluasi dirangkum pada Tabel 9.

Tabel 9. Hasil Evaluasi BERTScore

Metrik	Skor
Precision	0.79
Recall	0.83
F1 Score	0.81

Skor F1 sebesar 0.81 pada BERTScore menunjukkan bahwa model tidak hanya mampu menyalin kata-kata yang mirip secara permukaan, tetapi juga memahami makna umum dari teks sumber dan mengekspresikannya dalam bentuk ringkasan baru yang tetap kontekstual. Precision yang tinggi (0.79) mengindikasikan bahwa token dalam ringkasan model umumnya relevan terhadap referensi manual, sedangkan recall yang lebih tinggi (0.83) menunjukkan bahwa sebagian besar informasi penting berhasil dicakup dalam ringkasan yang dihasilkan.

Capaian ini sangat penting dalam konteks *abstractive summarization*, karena BERTScore

mengukur *semantic similarity* berdasarkan representasi *embedding* dari token, bukan hanya kecocokan bentuk. Skor yang tinggi mencerminkan keberhasilan model dalam menangkap relasi antar-entitas dalam kalimat serta kemampuan generalisasi semantik pada data dengan variasi struktur naratif dan teknis yang tinggi.

### 3.3.3 Evaluasi Token-Level

Untuk mengukur ketepatan model dalam mempertahankan elemen domain yang bersifat kritis seperti lokasi, nama paket, volume fisik, dan nilai kontrak, dilakukan evaluasi berbasis token-level terhadap entitas tersebut. Hasil evaluasinya dirangkum pada Tabel 10.

Tabel 10. Evaluasi Token Level

Metrik	Skor
Precision	0.77
Recall	0.69
F1 Score	0.73
Accuracy	0.71

Evaluasi token-level menjadi sangat penting dalam aplikasi nyata, karena akurasi informasi yang ditampilkan dalam ringkasan harus tetap mencerminkan fakta proyek dengan benar. Precision sebesar 0.77 menandakan bahwa sebagian besar entitas yang dihasilkan dalam ringkasan adalah valid dan sesuai dengan referensi. Sementara itu, recall sebesar 0.69 menunjukkan adanya ruang perbaikan dalam hal cakupan informasi, yang mungkin terlewat dalam beberapa kasus, terutama pada teks dengan struktur yang tidak baku.

Akurasi keseluruhan sebesar 0.71 merupakan indikasi bahwa model sudah cukup andal untuk diterapkan dalam sistem notifikasi atau pelaporan proyek secara otomatis. Capaian ini juga mengonfirmasi bahwa strategi pelatihan tokenizer khusus memainkan peran penting dalam membekali model dengan kemampuan memahami struktur frasa teknis khas domain pemerintahan.

Hasil evaluasi kuantitatif ini memperkuat argumen bahwa model BART yang dibangun dari nol tetap memiliki potensi besar untuk digunakan dalam domain sempit berbahasa Indonesia. Skor-skor yang diperoleh tidak hanya kompetitif dibandingkan baseline konvensional, tetapi juga menunjukkan bahwa efisiensi arsitektur ringan dapat sejalan dengan efektivitas hasil, terutama jika dikombinasikan dengan pendekatan *domain-aware tokenization* dan desain dataset yang representatif.

### 3.4 Ringkasan Otomatis

Untuk memberikan gambaran konkret mengenai kemampuan model dalam menyederhanakan informasi semi-struktural, salah satu contoh keluaran dari sistem terhadap entri data proyek aktual dirangkum pada Tabel 11.

Tabel 11. Contoh Teks Input &amp; Output Ringkasan Model

Teks Input:	Output Ringkasan (Model):
“Pekerjaan rehabilitasi saluran primer Desa Maju sepanjang 1,8 km dengan biaya 2,4 miliar rupiah untuk mendukung pengairan lahan sawah 150 hektar.”	“Rehabilitasi saluran primer Desa Maju 1,8 km biaya 2,4 miliar mendukung 150 hektar.”

Ringkasan yang dihasilkan menunjukkan bahwa model BART mampu mereduksi kalimat kompleks menjadi bentuk padat dan terstruktur tanpa kehilangan informasi inti. Unsur penting seperti jenis kegiatan, lokasi, ukuran fisik (1,8 km), nilai anggaran (2,4 miliar rupiah), dan tujuan proyek (pengairan 150 hektar) berhasil dipertahankan secara utuh. Kalimat yang dihasilkan meskipun bersifat *telegraphic* (ringkas tanpa konjungsi), tetap menyampaikan makna utama dan cocok untuk konteks presentasi informasi yang bersifat *machine-readable* seperti tampilan notifikasi otomatis atau widget dashboard proyek.

Fleksibilitas model dalam mempersingkat deskripsi panjang menjadi representasi yang esensial, namun tetap akurat secara semantik, menjadi nilai tambah dalam sistem yang ditujukan untuk efisiensi informasi publik. Hal ini relevan dalam implementasi sektor pemerintahan, di mana ringkasan digunakan sebagai pengingat keputusan atau bahan pelaporan internal yang cepat dan mudah dicerna oleh pemangku kebijakan.

### 3.5 Pembahasan Kinerja Model

Berdasarkan hasil evaluasi dan keluaran aktual model, dapat disimpulkan bahwa model BART yang dibangun dari nol menunjukkan performa yang kompeten dan adaptif dalam menghadapi tantangan data semi-struktural dalam domain proyek pemerintahan. Beberapa poin penting yang dapat ditarik dari hasil eksperimen adalah sebagai berikut:

#### 1. Stabilitas Pelatihan

Konvergensi nilai *training loss* dari 4.2 ke 1.7 selama lima epoch tanpa fluktuasi ekstrem menunjukkan bahwa arsitektur model yang ringan (hanya dua layer) cukup efektif dalam menyerap pola semantik dari dataset terbatas. Tidak ditemukannya gejala *overfitting* menandakan bahwa desain parameter yang konservatif justru menjadi keunggulan dalam domain sempit, karena mencegah model mempelajari noise dari teks teknis yang tidak terstandar.

#### 2. Adaptasi terhadap Domain

Kemampuan model untuk mempertahankan entitas teknis seperti lokasi proyek, nilai kontrak, dan ukuran fisik dapat dikaitkan dengan pelatihan tokenizer yang dirancang khusus berdasarkan korpus domain SIPAKAT AIR. Strategi ini menunjukkan bahwa model BART dari nol tetap dapat mengembangkan

pemahaman semantik domain-spesifik jika dibekali segmentasi token yang sesuai. Hal ini mendukung klaim bahwa *domain adaptation* tidak selalu memerlukan *pretrained model*, tetapi dapat dicapai melalui desain pipeline pelatihan yang terintegrasi.

### 3. Performa Metrik

Skor ROUGE-1 dan ROUGE-L yang melampaui 0.50 serta BERTScore F1 sebesar 0.81 mengonfirmasi bahwa ringkasan hasil model mendekati ringkasan referensi, baik secara leksikal maupun semantik. Kinerja token-level dengan akurasi 0.71 memperkuat temuan bahwa model berhasil mempertahankan elemen faktual proyek secara tepat, menjadikannya layak untuk diintegrasikan ke sistem pelaporan atau notifikasi publik. Konsistensi hasil evaluasi ini memperlihatkan bahwa performa model tidak hanya baik dalam satu metrik, tetapi seimbang dalam berbagai dimensi kualitas.

### 4. Keterbatasan dan Tantangan

Meskipun hasil cukup menjanjikan, model menunjukkan kecenderungan untuk menyederhanakan teks secara agresif pada masukan yang lebih panjang atau memiliki struktur sintaksis tidak baku. Dalam beberapa kasus, hal ini menyebabkan hilangnya detail penting seperti nama instansi pelaksana atau kerangka waktu proyek. Tantangan ini menunjukkan perlunya pendekatan *post-editing* atau *validation rule* tambahan untuk meningkatkan keterbacaan dan presisi semantik. Selain itu, skala dataset yang masih terbatas (hanya 200 pasang data) menjadi hambatan dalam memperluas cakupan linguistik model terhadap variasi kalimat proyek yang lebih kompleks.

Secara keseluruhan, temuan dari pembahasan ini menegaskan bahwa dengan pendekatan yang tepat—yakni pelatihan dari nol, arsitektur efisien, dan desain tokenizer domain-aware—pengembangan sistem ringkasan otomatis untuk aplikasi pemerintahan dapat dilakukan secara ringan namun tetap menghasilkan performa yang aplikatif. Model ini membuka ruang untuk riset lanjut di bidang *low-resource summarization*, terutama dalam konteks pengembangan NLP berbahasa Indonesia.

#### 3.6 Analisis Kesalahan dan Keterbatasan

Meskipun model menunjukkan performa yang kompetitif dalam evaluasi kuantitatif dan kualitas ringkasan otomatis, sejumlah kelemahan masih teridentifikasi yang perlu dicermati untuk pengembangan di tahap selanjutnya. Analisis kesalahan ini dilakukan untuk memahami keterbatasan model dalam konteks struktur arsitektural, representasi data, dan konteks aplikatif.

##### 1. Underfitting Arsitektural

Arsitektur model yang hanya terdiri dari dua layer encoder-decoder dengan hidden size sebesar 256 kemungkinan belum cukup dalam untuk menangkap konteks semantik yang kompleks dan panjang. Meskipun desain ini menghindari risiko overfitting dan mengurangi beban komputasi, kemampuan representasional model terhadap sintaks dan dependensi semantik yang lebih dalam menjadi terbatas. Hal ini menyebabkan ringkasan kadang kehilangan konteks naratif yang lebih luas, terutama pada entri proyek yang mengandung deskripsi berlapis.

##### 2. Reduksi Berlebihan

Model menunjukkan kecenderungan untuk menyederhanakan kalimat dengan menghilangkan unsur penting, seperti satuan teknis (misalnya "km", "hektar"), jenis pekerjaan spesifik (seperti "normalisasi" atau "peningkatan kapasitas saluran"), serta rincian penunjang seperti tujuan proyek. Meskipun strategi ini menjaga ringkasan tetap singkat, hilangnya informasi krusial dapat mengurangi validitas ringkasan dalam konteks pelaporan atau notifikasi berbasis data teknis. Fenomena ini menunjukkan pentingnya kontrol terhadap *semantic fidelity* dan kebutuhan akan mekanisme *post-editing* atau *constraint-based decoding* di masa depan.

##### 3. Keterbatasan Data

Dataset yang digunakan dalam penelitian ini relatif kecil (200 pasangan teks-ringkasan), dan meskipun data berasal dari sumber riil, variasi struktur kalimat dan tipe proyek masih terbatas. Hal ini berdampak pada generalisasi model terhadap bentuk kalimat yang lebih kompleks atau tidak baku. Sebagai akibatnya, model bisa gagal menangani kasus-kasus *out-of-distribution* (OOD) yang tidak terwakili dalam data pelatihan. Penambahan jumlah data, serta diversifikasi format input dan gaya bahasa, akan sangat meningkatkan ketahanan model terhadap variasi masukan.

Untuk penelitian selanjutnya, disarankan dilakukan visualisasi kesalahan berdasarkan tipe kesalahan linguistik (misalnya: hilangnya subjek, kesalahan numerik, kekurangan makna), serta pembedaan manual (error tagging) untuk memperkuat umpan balik ke dalam sistem pelatihan. Teknik *data augmentation*, *curriculum learning*, atau penggunaan *reinforcement learning from human feedback (RLHF)* juga dapat menjadi alternatif untuk peningkatan ketepatan semantik model.

#### 3.7 Analisis Generalisasi Model

Berdasarkan perbandingan antara training loss dan validation loss, tidak ditemukan indikasi

overfitting. Perbedaan performa antara validation dan testing juga sangat kecil, yang menunjukkan bahwa model mampu melakukan generalisasi dengan baik meskipun ukuran dataset terbatas. Hal ini dapat terjadi karena arsitektur ringan dengan dua layer mampu meminimalkan risiko mempelajari noise pada data.

### 3.8 Perbandingan dengan Studi Sebelumnya

Untuk menempatkan hasil penelitian dalam konteks yang lebih luas, model BART dari nol yang dikembangkan dalam studi ini dibandingkan dengan beberapa pendekatan sebelumnya yang memanfaatkan model pretrained berskala besar, seperti IndoBART dan Pegasus. Perbandingan dilakukan berdasarkan tiga metrik evaluasi utama, sebagaimana dirangkum dalam Tabel 12.

Tabel 12. Perbandingan Kinerja dengan Studi Sebelumnya

Model	ROUGE-I F1	ROUGE-L F1	BERT Score F1	Catatan
IndoBART (berita)	0.55	0.53	0.85	Pretrained, domain umum
Pegasus (medis)	0.58	0.54	0.83	Pretrained besar, domain medis
BART (dari nol)	0.5080	0.5082	0.81	Ringan, domain sempit, tidak pretrained

Dari tabel tersebut, terlihat bahwa meskipun model yang dikembangkan dalam penelitian ini tidak mengungguli model pretrained dalam metrik absolut, perbedaannya tidak signifikan untuk ukuran domain teknis dengan data terbatas. Justru, pendekatan *from scratch* menawarkan sejumlah keunggulan penting:

- Fleksibilitas Domain**  
Model yang dibangun dari awal dapat dikustomisasi dengan struktur arsitektural dan tokenizer yang disesuaikan secara spesifik terhadap karakteristik data domain SIPAKAT AIR, menjadikannya lebih adaptif dalam mengenali struktur linguistik dan entitas lokal.
- Efisiensi Komputasi**  
Model ringan ini lebih hemat sumber daya, cocok untuk pengembangan sistem NLP di lingkungan dengan keterbatasan infrastruktur komputasi seperti instansi pemerintah daerah atau platform berbasis cloud ringan.
- Kemandirian Infrastruktur Bahasa**  
Dalam konteks pengembangan NLP berbahasa Indonesia, pendekatan ini menunjukkan bahwa ketergantungan pada pretrained model global (yang sering kali kurang relevan terhadap struktur kalimat lokal) dapat dikurangi melalui desain pipeline yang terfokus pada domain.

Dengan demikian, perbandingan ini menunjukkan bahwa model ringan dan dilatih dari awal tetap kompetitif dan relevan untuk digunakan dalam skenario praktis, terutama untuk tugas *domain-specific summarization* di sektor publik. Penelitian ini mengisi celah literatur yang masih langka dalam pemodelan teks semi-struktural di lingkungan berbahasa Indonesia, sekaligus membuka jalan untuk eksplorasi lebih lanjut terhadap efisiensi dan efektivitas pelatihan lokal berbasis NLP.

### 3.9 Rencana Implementasi Sistem

Model dan tokenizer hasil pelatihan telah disimpan secara terstruktur dalam format yang kompatibel dengan ekosistem Python dan Huggingface, dan kini siap untuk diintegrasikan ke dalam sistem SIPAKAT AIR. Proses integrasi dirancang melalui antarmuka layanan berbasis REST API, yang memungkinkan pemanggilan model secara efisien dan fleksibel dari berbagai komponen frontend atau aplikasi pihak ketiga. Secara umum, alur rencana implementasi dijelaskan pada Tabel 13.

Tabel 13. Rencana Alur Implementasi Sistem

Alur	Keterangan
<b>Input</b>	Teks uraian proyek infrastruktur yang diambil langsung dari antarmuka pengguna SIPAKAT AIR.
<b>Proses</b>	Permintaan dikirim ke server backend melalui API yang mengeksekusi model summarization untuk menghasilkan ringkasan berbasis input tersebut.
<b>Output</b>	Hasil ringkasan dikembalikan dalam format terstruktur seperti JSON, sehingga mudah untuk diolah lebih lanjut.
<b>Tampilan</b>	Ringkasan otomatis dapat ditampilkan pada dashboard proyek, ditanamkan ke dalam notifikasi berkala (email, SMS, atau push notification), atau dijadikan elemen dalam laporan teknis berformat (misalnya PDF atau rekap digital).

Dalam implementasi teknisnya, sistem direncanakan menggunakan framework FastAPI karena keunggulannya dalam performa dan kemudahan pengembangan layanan REST modern, meskipun alternatif seperti Flask juga disiapkan sebagai fallback. Selain itu, sistem dapat diperluas untuk mendukung pengarsipan data ke basis data SQL/NoSQL, integrasi dengan sistem monitoring proyek, serta fungsionalitas tambahan seperti autentikasi pengguna, audit trail, dan logging untuk kebutuhan forensik data.

Dengan pendekatan ini, proses ringkasan proyek yang sebelumnya memerlukan intervensi manual dapat dilakukan secara otomatis, akurat, dan real-time, sehingga mendukung efisiensi pelaporan dan peningkatan transparansi dalam pengelolaan infrastruktur publik.

### 3.10 Implikasi Temuan

Hasil dari studi ini memberikan kontribusi penting baik dari sisi akademik maupun penerapan praktis. Beberapa implikasi utama yang dapat disorot antara lain:

1. Pengembangan model ringan yang dibangun dari nol (tanpa bobot pra-latih) terbukti tetap mampu memberikan performa yang kompetitif untuk domain sempit, seperti proyek infrastruktur sumber daya air. Temuan ini menjadi validasi bahwa model from scratch masih relevan, terutama di lingkungan dengan keterbatasan komputasi dan data, seperti pada instansi pemerintah atau sistem berbasis lokal.
2. Penggunaan tokenizer domain-spesifik yang dilatih menggunakan korpus internal terbukti secara signifikan meningkatkan akurasi pemrosesan token, representasi semantik, dan kemampuan model dalam mengenali serta mempertahankan entitas kritis seperti lokasi, volume pekerjaan, atau nilai kontrak. Hal ini menunjukkan pentingnya menyesuaikan komponen NLP terhadap struktur dan karakteristik data yang khas dari domain tertentu.
3. Evaluasi yang bersifat multi-dimensi—meliputi ROUGE untuk kesesuaian leksikal, BERTScore untuk pemahaman semantik, evaluasi token-level untuk akurasi entitas memberikan gambaran menyeluruh terhadap kinerja model. Strategi evaluatif ini dapat menjadi praktik terbaik (best practice) dalam pengembangan sistem NLP yang ditujukan untuk aplikasi nyata.

Dari sisi aplikatif, sistem summarization otomatis yang dikembangkan tidak hanya memiliki nilai tambah dalam konteks SIPAKAT AIR, tetapi juga dapat menjadi template atau blueprint untuk pengembangan sistem informasi serupa di sektor publik lainnya, seperti perencanaan anggaran, pengadaan barang dan jasa, atau pelaporan kegiatan pembangunan. Integrasi teknologi NLP seperti ini memiliki potensi besar untuk mendorong transformasi digital di pemerintahan, mempercepat diseminasi informasi teknis, serta meningkatkan keterbukaan dan akuntabilitas kepada publik.

Secara keseluruhan, penelitian ini tidak hanya menunjukkan kelayakan teknis pengembangan sistem summarization domain sempit berbasis BART dari awal, tetapi juga menggariskan pentingnya pendekatan berbasis data lokal, pemodelan ringan, dan evaluasi kontekstual dalam pengembangan solusi NLP yang berorientasi pada kebutuhan masyarakat dan sektor publik Indonesia.

#### 4. Kesimpulan

Penelitian ini menunjukkan bahwa model *abstractive summarization* berbasis BART yang dibangun dari awal mampu memberikan performa yang kompetitif dalam meringkas teks semi-struktural pada domain terbatas, khususnya data proyek infrastruktur dari sistem SIPAKAT AIR.

Dengan arsitektur *encoder-decoder* ringan dan *tokenizer* khusus yang dilatih dari korpus internal, model ini menunjukkan performa yang menjanjikan dan berpotensi kompetitif apabila dibandingkan dengan baseline sederhana. Namun, efektivitasnya pada skala data yang lebih besar masih perlu diuji. Evaluasi kuantitatif menunjukkan pencapaian skor ROUGE-1 dan ROUGE-L di atas 0,50 serta BERTScore F1 sebesar 0,81, yang menandakan bahwa ringkasan yang dihasilkan mendekati kualitas referensi manual secara semantik. Temuan ini menegaskan bahwa meskipun tanpa bobot *pretrained*, model tetap mampu menghasilkan ringkasan yang padat, relevan, dan kontekstual. Namun, studi ini masih bersifat eksploratif dan belum diuji dalam lingkungan operasional secara langsung, sehingga efektivitasnya dalam konteks *real-time* perlu dikaji lebih lanjut. Keterbatasan lainnya mencakup ukuran dataset yang terbatas serta belum dilibatkannya pengguna akhir dalam proses evaluasi, yang dapat memengaruhi generalisasi model terhadap variasi teks yang lebih kompleks. Untuk penelitian lebih lanjut disarankan untuk memperluas korpus data, mengembangkan ringkasan manual yang lebih representatif, serta membangun REST API yang memungkinkan integrasi langsung dengan sistem SIPAKAT AIR. Implikasi dari penelitian ini melampaui konteks SIPAKAT AIR, karena pendekatan *from-scratch* dan penggunaan *tokenizer* domain-spesifik terbukti dapat meningkatkan relevansi dan akurasi model dalam menangani teks teknis berbahasa Indonesia. Secara keseluruhan, studi ini memberikan kontribusi metodologis bagi pengembangan NLP lokal dan membuka peluang pemanfaatan *summarization* otomatis dalam sistem digital pemerintahan yang transparan, efisien, dan berbasis data.

#### Daftar Pustaka:

- Huang, D., Cui, L., Yang, S., Bao, G., Wang, K., Xie, J., & Zhang, Y. (2020). *What Have We Achieved on Text Summarization?* <https://github.com/hddbang/PolyTope>
- Irsan, I. C., Zhang, T., Thung, F., Lo, D., & Jiang, L. (2022). *AutoPRTtitle: A Tool for Automatic Pull Request Title Generation.* <http://arxiv.org/abs/2206.11619>
- Jearanaitanakij, K., Boonpong, S., Teainnagrm, K., Thonglor, T., Kullawan, T., & Yongpiyakul, C. (2024). Fast Hybrid Approach for Thai News Summarization. *Technol. Horiz.*, 41(4), 410307. <https://doi.org/10.55003/ETH.410307>
- Ketimeni, S., & Sheela, J. (2024). Modified CNN with Transfer Learning for Multi-Document Summarization: Proposing Co-Occurrence Matrix Generation-Based Knowledge Extraction. *International Journal of Uncertainty, Fuzziness and Knowledge-Based*

- Systems, 32(05), 721–745.  
<https://doi.org/10.1142/S021848852450017X>
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., & Ai, F. (n.d.). *BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension*.  
<https://huggingface.co/transformers>
- Liu, Y., & Lapata, M. (2019). *Text Summarization with Pretrained Encoders*. <https://github.com/>
- Mehamed, M. A., Xiong, S., & Aberha, A. F. (2025). Hybrid Approach for Automatic Text Summarization for Low-resourced Amharic Language. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 24(7).  
<https://doi.org/10.1145/3743677>
- Mehta, V., Patel, K., Shah, A., & Thakkar, S. (2022). A Survey on Automatic Text Summarization. *International Research Journal of Engineering and Technology*. [www.irjet.net](http://www.irjet.net)
- Prakash, A. (2025). Enhancing News Article Summarization with Machine Learning. *International Journal for Global Academic & Scientific Research*, 3(4), 20–34.  
<https://doi.org/10.55938/ijgasr.v3i4.152>
- Saeed, M. Y., Awais, M., Younas, M., Shah, M. A., Khan, A., Uddin, M. I., & Mahmoud, M. (2021). An abstractive summarization technique with variable length keywords as per document diversity. *Computers, Materials and Continua*, 66(3), 2409–2423.  
<https://doi.org/10.32604/cmc.2021.014330>
- Suliman, W., Yaseen, A., & Hamada, N. (2025). Advancements in abstractive text summarization: a deep learning approach. *IAES International Journal of Artificial Intelligence*, 14(3), 2315–2327.  
<https://doi.org/10.11591/ijai.v14.i3.pp2315-2327>
- V, N., & R, M. T. (n.d.). Improved Bi-GRU framework for Multi-document text summarization with aspect and thematic feature descriptor: Model training via hybrid optimization. *Intelligent Decision Technologies*, 0(0), 18724981251346852.  
<https://doi.org/10.1177/18724981251346852>
- Wiratmoko, G., Thamrin, H., & Pamungkas, E. W. (2025). Performance of Machine Learning Algorithms on Automatic Summarization of Indonesian Language Texts. *Jurnal Online Informatika*, 10(1), 196–204.  
<https://doi.org/10.15575/join.v10i1.1506>
- Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2020). *PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization*.  
<http://arxiv.org/abs/1912.08777>
- Zhang, M., Zhou, G., Yu, W., Huang, N., & Liu, W. (2022). A Comprehensive Survey of Abstractive Text Summarization Based on Deep Learning. In *Computational Intelligence and Neuroscience* (Vol. 2022). Hindawi Limited.  
<https://doi.org/10.1155/2022/7132226>
- Zin, M. M., Nguyen, H. T., Satoh, K., Sugawara, S., & Nishino, F. (2023). Information Extraction from Lengthy Legal Contracts: Leveraging Query-Based Summarization and GPT-3.5. *Frontiers in Artificial Intelligence and Applications*, 379, 177–186.  
<https://doi.org/10.3233/FAIA230963>

*Halaman ini sengaja dikosongkan*