PERANCANGAN DAN EVALUASI SISTEM INFORMASI PERINGKASAN OTOMATIS DOKUMEN POK KEUANGAN BERBASIS BART

DESIGN AND EVALUATION OF AN INFORMATION SYSTEM FOR AUTOMATIC SUMMARIZATION OF FINANCIAL POK DOCUMENTS USING THE BART MODEL

Sofwatul Ummah[1], Supriyono[2*], Okta Qomaruddin Aziz[3]

230605110040@student.uin-malang.ac.id[1], priyono@ti.uin-malang.ac.id[2*]
okta.qomaruddin@uin-malang.ac.id[3]

[1,2,3]Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang

Abstrak

Peningkatan volume dokumen keuangan pemerintah daerah menyebabkan proses pencarian informasi menjadi semakin kompleks, khususnya pada Dokumen Petunjuk Operasional Kegiatan (POK) yang memiliki struktur semi-terstruktur dan kepadatan entitas numerik tinggi. Penelitian ini bertujuan merancang dan mengevaluasi sistem informasi peringkasan otomatis dokumen POK keuangan menggunakan model *Bidirectional* and *Auto-Regressive Transformers* (BART). Tahapan penelitian meliputi *preprocessing* dokumen, tokenisasi, penerapan mekanisme *chunking* untuk menangani dokumen panjang, *fine-tuning model,* serta evaluasi menggunakan metrik ROUGE dan analisis inferensi. Dataset penelitian terdiri dari dokumen POK keuangan pemerintah daerah berbahasa Indonesia yang dikonversi menjadi pasangan document-summary. Hasil pengujian menunjukkan bahwa model memperoleh ROUGE-1 sebesar 0,8971, ROUGE-2 sebesar 0,8916, ROUGE-L sebesar 0,9021, dan ROUGE-Lsum sebesar 0,9008 dengan rata-rata latensi inferensi sekitar 0,612 detik per dokumen. Selain itu, mekanisme *chunking mampu* mempertahankan kualitas ringkasan tanpa memberikan pengaruh signifikan terhadap performa inferensi. Penelitian ini juga menunjukkan bahwa evaluasi berbasis ROUGE pada dokumen finansial memiliki keterbatasan dalam menangkap ketepatan informasi numerik secara kontekstual. Hasil penelitian menunjukkan bahwa model BART memiliki potensi untuk diterapkan pada sistem pengelolaan dokumen keuangan pemerintah daerah.

Kata kunci*: *Sistem Informasi, Peringkasan Otomatis, BART, Pemrosesan Bahasa Alami, Dokumen POK, Keuangan.

Abstract

The increasing volume of regional government financial documents has made information retrieval more complex, particularly for Operational Activity Instruction (POK) documents that contain semi-structured content and dense numerical entities. This study aims to design and evaluate an automatic summarization information system for financial POK documents using the Bidirectional and Auto-Regressive Transformers (BART) model. The research stages included document preprocessing, tokenization, implementation of a chunking mechanism for handling long documents, model fine-tuning, and evaluation using ROUGE metrics and inference analysis. The dataset consisted of Indonesian regional government financial POK documents converted into document-summary pairs. Experimental results show that the model achieved a ROUGE-1 score of 0.8971, ROUGE-2 of 0.8916, ROUGE-L of 0.9021, and ROUGE-Lsum of 0.9008, with an average inference latency of approximately 0.612 seconds per document. In addition, the chunking mechanism was able to maintain summarization quality without significantly affecting inference performance. The study also found that ROUGE-based evaluation on financial

documents has limitations in capturing the contextual accuracy of numerical information. Overall, the findings indicate that the BART model has potential for implementation in regional government financial document management systems.

***Keywords**: Information Systems, Automatic Summarization, BART, Natural Language Processing , POK Documents, Finance.*

1. PENDAHULUAN

Transformasi digital pada sektor pemerintahan meningkatkan volume dokumen administratif dan keuangan yang harus diproses secara cepat dan akurat. Salah satu dokumen penting dalam pengelolaan anggaran pemerintah daerah adalah Dokumen Petunjuk Operasional Kegiatan (POK), yang memuat informasi program, kode wilayah, serta rincian anggaran. Namun, struktur dokumen yang semi-terstruktur, repetitif, dan kaya entitas numerik menyebabkan proses identifikasi informasi penting menjadi tidak efisien apabila dilakukan secara manual. Kondisi ini mendorong kebutuhan terhadap sistem peringkasan otomatis yang mampu membantu pengguna memahami isi dokumen secara lebih cepat.

Automatic Text Summarization (ATS) merupakan salah satu pendekatan *Natural Language Processing* (NLP) untuk menghasilkan ringkasan dari dokumen panjang tanpa menghilangkan informasi utama. Perkembangan model transformer seperti BERT, T5, PEGASUS, dan BART telah meningkatkan performa *abstractive summarization* pada berbagai domain [1]–[5]. Di antara model tersebut, *Bidirectional and Auto-Regressive Transformers* (BART) menunjukkan performa yang kompetitif karena menggabungkan *encoder Bidirectional* dan d*ecoder autoregressive* dalam arsitektur *sequence-to-sequence* [6]. Beberapa penelitian juga menunjukkan bahwa BART efektif digunakan pada summarization dokumen panjang melalui penerapan mekanisme *chunking*[7], [8].

Meskipun penelitian summarization telah berkembang pesat, sebagian besar studi masih berfokus pada domain berita, artikel ilmiah, dan dokumen umum [1], [9]. Penelitian pada dokumen keuangan pemerintah berbahasa Indonesia masih terbatas, padahal dokumen POK memiliki karakteristik yang berbeda dari teks naratif biasa, terutama pada dominasi entitas numerik dan pola administratif yang berulang. Selain itu, penelitian sebelumnya umumnya hanya berfokus pada skor evaluasi agregat seperti ROUGE tanpa membahas distribusi error, perilaku model, maupun keterbatasan evaluasi pada domain numerik-intensif [10].

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan sistem informasi peringkasan otomatis dokumen POK keuangan berbasis model BART dengan mekanisme *chunking* untuk menangani dokumen panjang. Penelitian tidak hanya mengevaluasi kualitas ringkasan menggunakan ROUGE, tetapi juga menganalisis distribusi skor per-sampel, latensi inferensi, compression ratio, serta *error analysis* pada kasus keberhasilan dan kegagalan model. Tingginya skor ROUGE pada penelitian ini tidak sepenuhnya merepresentasikan kemampuan abstraksi semantik yang kompleks, melainkan juga dipengaruhi oleh karakteristik dokumen POK yang memiliki pola linguistik administratif yang repetitif dan semi-terstruktur. Oleh karena itu, performa model perlu diinterpretasikan dalam konteks domain-specific summarization, bukan sebagai generalisasi terhadap seluruh jenis dokumen bahasa Indonesia.

Kontribusi utama penelitian ini meliputi: (1) adaptasi model BART pada dokumen POK keuangan berbahasa Indonesia, (2) implementasi mekanisme *chunking* untuk mengatasi keterbatasan panjang token, serta (3) evaluasi multidimensional terhadap kualitas ringkasan dan perilaku inferensi model pada domain dokumen keuangan pemerintah daerah.

2. METODOLOGI

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan experimental research berbasis *Natural Language Processing* (NLP) untuk membangun sistem peringkasan otomatis dokumen Petunjuk Operasional Kegiatan (POK) keuangan menggunakan model *Bidirectional* and *Auto-Regressive Transformers* (BART). Pipeline penelitian terdiri atas enam tahap utama, yaitu dataset construction, preprocessing, tokenization, *chunking*, *fine-tuning* model, dan evaluasi performa. Seluruh eksperimen diimplementasikan menggunakan framework *Hugging Face Transformers* berbasis PyTorch.

2.2 Dataset Penelitian

Dataset penelitian terdiri dari dokumen POK keuangan pemerintah daerah di Provinsi Jawa Timur yang memuat informasi administratif dan finansial seperti kode wilayah, nama kabupaten/kota, program kegiatan, serta nominal anggaran dalam format Rupiah. Dokumen dikonversi menjadi pasangan document-summary untuk kebutuhan supervised learning. Karakteristik dataset meliputi struktur semi-terstruktur, dominasi entitas numerik, serta panjang dokumen yang bervariasi.

2.3 *preprocessing* Data

Tahap *preprocessing* dilakukan untuk meningkatkan konsistensi representasi teks sebelum pelatihan model. Proses ini mencakup pembersihan karakter non-UTF, normalisasi spasi dan simbol, penghapusan artefak ekstraksi dokumen, serta normalisasi format nominal Rupiah. Karena sebagian dokumen berasal dari struktur tabular, dilakukan konversi tabel ke format naratif menggunakan pendekatan table-to-text transformation. Selain itu, entitas numerik finansial dipertahankan secara eksplisit untuk mengurangi hilangnya informasi penting selama summarization. Dataset kemudian divalidasi menggunakan quality control berbasis panjang token minimum, sinkronisasi dokumen dan ringkasan, serta pengecekan duplikasi data.

2.4 Tokenization dan Representasi Teks

Representasi teks dilakukan menggunakan tokenizer bawaan model *facebook/bart-large-cnn* dari *HuggingFace Transformers* berbasis *Byte-Pair Encoding* (BPE). Setiap dokumen direpresentasikan dalam bentuk input_ids, attention_mask, dan labels. Karena sebagian dokumen melebihi batas konteks model, diterapkan mekanisme *chunking* dengan membagi dokumen menjadi beberapa segmen berdasarkan panjang token maksimum. Setiap chunk diproses secara independen, kemudian hasil ringkasan digabungkan menjadi output akhir. Pendekatan ini digunakan untuk menjaga stabilitas memori dan memungkinkan pemrosesan dokumen panjang tanpa memotong informasi utama.

2.5 *fine-tuning* Model

Penelitian menggunakan model *abstractive summarization* BART varian *facebook/bart-large-cnn* yang di-fine-tune menggunakan framework PyTorch dan *HuggingFace Trainer*. Proses pelatihan dilakukan selama 3 *epoch* menggunakan optimizer AdamW dengan learning rate 2×10−5.Objective training menggunakan cross−entropy sequence loss:

$L = -\sum_{t=1}^{T} \log P(y_t | y_{<t}, X)$ ..... (1)

Selama inferensi, model menghasilkan ringkasan secara autoregressive berdasarkan representasi *encoder-decoder transformer*.

2.6 Evaluasi Model

Evaluasi dilakukan menggunakan metrik ROUGE yang terdiri dari ROUGE-1, ROUGE-2, ROUGE-L, dan ROUGE-Lsum untuk mengukur kesamaan antara ringkasan hasil model dan referensi. Formula evaluasi ROUGE direpresentasikan sebagai:

$$ROUGE - N = \frac{\sum_{\in\{} \quad \}\ \sum_{gram_n \in} Count_{match}(gram_n)}{\sum_{\in\{} \quad \}\ \sum_{gram_n \in} Count(gram_n)} \quad \ldots\ldots (2)$$

Selain evaluasi kualitas ringkasan, penelitian juga mengukur compression ratio, distribusi ROUGE per-sampel, dan latensi inferensi. Compression ratio dihitung menggunakan perbandingan panjang ringkasan terhadap panjang dokumen:

$$Compression\ Ratio = \frac{Length_{summary}}{Length_{document}} \quad \ldots\ldots (3)$$

Analisis tambahan dilakukan melalui *correlation heatmap* dan *error analysis* untuk mengidentifikasi hubungan antar-variabel seperti panjang dokumen, jumlah token output, latensi inferensi, serta missing entity numerik pada dokumen keuangan. Hasil evaluasi menunjukkan model memperoleh ROUGE-1 sebesar 0,8971, ROUGE-2 sebesar 0,8916, ROUGE-L sebesar 0,9021, dan ROUGE-Lsum sebesar 0,9008.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pelatihan Model

Proses pelatihan model dilakukan selama tiga *epoch* menggunakan arsitektur BART pada dokumen POK keuangan pemerintah daerah. Kurva pelatihan pada Gambar 1 menunjukkan penurunan training loss dari sekitar 9,9 pada tahap awal menjadi sekitar 1,2 pada akhir pelatihan. *Evaluation loss* juga stabil di bawah 0,30 setelah pertengahan proses training, yang menunjukkan bahwa model berhasil beradaptasi terhadap distribusi dokumen POK keuangan tanpa indikasi overfitting yang signifikan.

Gambar 1. Kurva Training Loss dan Validation Loss selama Proses *fine-tuning model* BART pada Dokumen POK Keuangan (3 Epoch, 45 Steps)

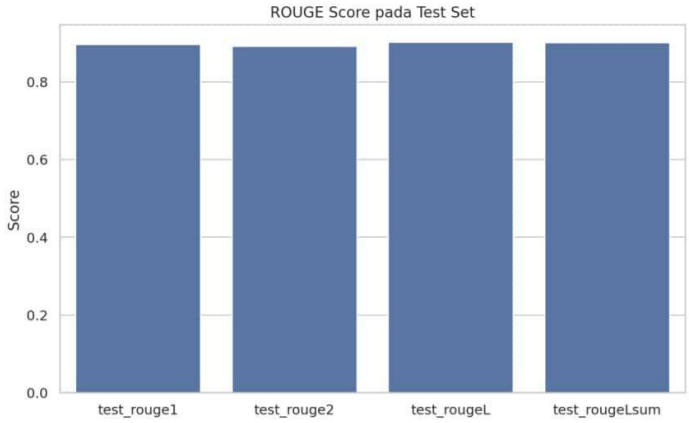
Perbedaan antara training loss dan *Evaluation loss* dipengaruhi oleh penggunaan *dropout* selama proses pelatihan serta ukuran dataset evaluasi yang relatif kecil. Secara umum, pola konvergensi menunjukkan bahwa *fine-tuning* selama tiga *epoch* sudah memadai untuk menangani karakteristik dokumen POK yang memiliki pola administratif berulang dan semi-terstruktur.

Table 1. Ringkasan Metrik Evaluasi pada *Test Set*

Metric	Score
ROUGE-1	0.8971
ROUGE-2	0.8916
ROUGE-L	0.9021
ROUGE-Lsum	0.9008
Test Loss	0.3909
Runtime (s)	10.3185

3.2 Analisis Skor ROUGE pada *Test Set*

Hasil evaluasi kuantitatif pada *Test Set* yang terangkum dalam Tabel 1 dan divisualisasikan pada Gambar 2 menampilkan performa yang sangat kompetitif.



Gambar 2. Visualisasi Skor ROUGE pada *Test Set*

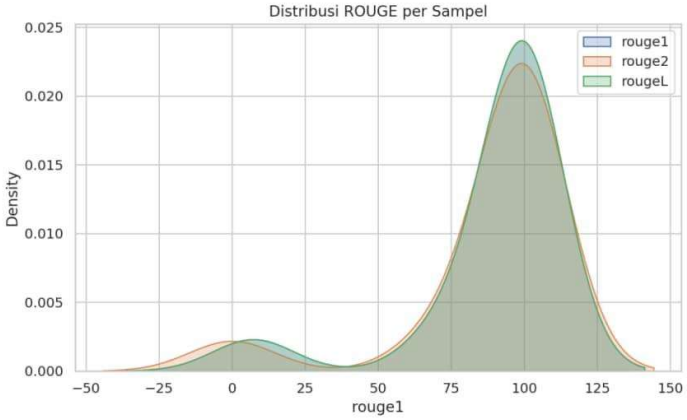
Model memperoleh ROUGE-1 sebesar 0,8971, ROUGE-2 sebesar 0,8916, ROUGE-L sebesar 0,9021, dan ROUGE-Lsum sebesar 0,9008. Nilai yang relatif konsisten pada seluruh metrik menunjukkan bahwa model mampu mempertahankan kesesuaian kata, frasa, dan urutan informasi dengan referensi ringkasan.

Perbedaan yang kecil antara ROUGE-1 dan ROUGE-2 mengindikasikan bahwa model tidak hanya mengenali entitas individual, tetapi juga mempertahankan hubungan kontekstual antarfrasa pada dokumen keuangan. Selain itu, nilai ROUGE-L yang tinggi menunjukkan bahwa struktur informasi pada ringkasan hasil prediksi tetap mengikuti pola urutan informasi pada dokumen referensi.

Distribusi skor per-sampel menunjukkan pola bimodal, di mana sebagian besar dokumen memperoleh skor tinggi, sementara sebagian kecil lainnya memiliki skor rendah. Hal ini menunjukkan bahwa performa model masih dipengaruhi oleh karakteristik tertentu dari dokumen input, khususnya pada dokumen dengan struktur dan distribusi entitas numerik yang berbeda dari data pelatihan.

Dibandingkan penelitian summarization bahasa Indonesia pada domain umum, performa model pada penelitian ini tergolong tinggi. Namun, tingginya skor ROUGE tidak sepenuhnya merepresentasikan kemampuan abstraksi semantik yang kompleks, melainkan juga dipengaruhi oleh karakteristik dokumen POK yang memiliki pola linguistik administratif yang repetitif dan semi-terstruktur. Oleh karena itu, performa model perlu diinterpretasikan dalam konteks domain-specific summarization, bukan sebagai generalisasi terhadap seluruh jenis dokumen bahasa Indonesia.

3.3 Analisis Distribusi ROUGE Per-Sampel



Gambar 3. Distribusi Skor ROUGE per Sampel pada *Test Set* Menunjukkan Pola Bimodal

Distribusi skor ROUGE per sampel pada Gambar 3 menunjukkan pola bimodal, dengan mayoritas sampel berada pada rentang skor tinggi (95–100) dan sebagian kecil lainnya berada pada skor rendah. Pola ini menunjukkan bahwa performa model tidak sepenuhnya seragam pada seluruh dokumen uji.

Kurva ROUGE-1, ROUGE-2, dan ROUGE-L yang hampir berimpit pada puncak distribusi mengindikasikan bahwa model mampu mempertahankan kesesuaian kata, frasa, dan struktur informasi secara konsisten pada sebagian besar sampel. Namun, terdapat beberapa kasus ekstrim dengan skor ROUGE-L di bawah 15 yang menunjukkan kegagalan model pada dokumen tertentu. Kasus ini mengindikasikan bahwa performa summarization masih dipengaruhi oleh karakteristik dokumen, terutama pada distribusi entitas numerik dan struktur administratif yang berbeda dari pola dominan data pelatihan.

Temuan ini menunjukkan bahwa skor agregat yang tinggi belum sepenuhnya merepresentasikan performa model pada seluruh sampel dokumen. Oleh karena itu, evaluasi per-sampel dan *error analysis* tetap diperlukan untuk mengidentifikasi potensi kegagalan model pada kondisi tertentu.

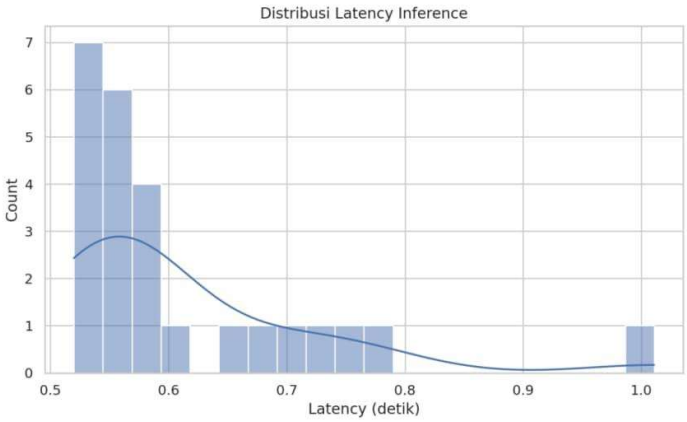
3.4 Analisis Compression Ratio dan Relasi dengan Kualitas Ringkasan

Hasil analisis menunjukkan bahwa mayoritas ringkasan yang dihasilkan model memiliki compression ratio pada rentang 0,83–0,88. Hal ini menunjukkan bahwa model cenderung mempertahankan sebagian besar informasi dari dokumen sumber dibandingkan melakukan kompresi agresif. Pada dokumen POK keuangan, perilaku tersebut masih relevan karena tingginya kepadatan entitas numerik dan administratif membuat penghilangan informasi berisiko mengurangi ketepatan isi ringkasan.

Beberapa sampel dengan compression ratio lebih rendah menunjukkan kualitas ringkasan yang lebih bervariasi. Kondisi ini mengindikasikan bahwa model masih mengalami kesulitan pada dokumen dengan pola struktur yang berbeda dari distribusi data pelatihan. Selain itu, terdapat kasus di mana ringkasan yang dihasilkan tetap faktual terhadap dokumen sumber namun memperoleh skor ROUGE rendah akibat perbedaan strategi summarization dengan referensi manusia.

Secara umum, model menunjukkan kecenderungan konservatif dalam melakukan kompresi teks. Dalam konteks dokumen keuangan pemerintah, pendekatan ini dinilai lebih sesuai karena ketepatan informasi numerik dan administratif lebih penting dibandingkan tingkat kompresi yang tinggi.

3.5 Analisis Latensi Inferensi



Gambar 4. Distribusi Latensi Inferensi Model BART per Dokumen POK Keuangan (dalam Detik)

Distribusi latensi inferensi pada Gambar 4 menunjukkan bahwa sebagian besar proses summarization diselesaikan dalam waktu kurang dari 0,65 detik. Berdasarkan Tabel 2, model memperoleh rata-rata latensi sebesar 0,612 detik dengan median 0,563 detik. Hasil ini menunjukkan bahwa model memiliki performa inferensi yang cukup stabil untuk pemrosesan dokumen POK keuangan.

Table 2. Tabel Statistik Deskriptif Latensi Inferensi

Statistik	Nilai (detik)	Statistik	Nilai (detik)
Mean	0.612	Min	0.520
Std Dev	0.112	Max	1.011
Median	0.563	Q3	0.646

Beberapa kasus outlier dengan latensi lebih tinggi berkaitan dengan panjang output ringkasan yang lebih besar. Temuan ini sejalan dengan hasil *correlation heatmap* yang menunjukkan hubungan positif antara jumlah token hasil generasi dan waktu inferensi. Sebaliknya, panjang dokumen input tidak menunjukkan peningkatan latensi secara linear karena dokumen panjang diproses menggunakan mekanisme *chunking*.

Secara umum, rata-rata latensi di bawah satu detik menunjukkan bahwa sistem masih layak digunakan untuk skenario semi-real-time pada pengelolaan dokumen keuangan pemerintah daerah.

3.6 *error analysis*: Kasus Kegagalan dan Penyebabnya

Error analysis dilakukan untuk mengidentifikasi pola kegagalan model yang tidak terlihat dari metrik agregat. Tabel 3 menunjukkan dua kasus dengan skor ROUGE terendah pada *Test Set*.

Table 3. Tabel Kasus Kegagalan Terburuk (*Worst Failure Cases*)

Idx	Doc Tokens	Gen Tokens	ROUGE-1	ROUGE-L	Latency (s)	Miss Rupiah
14	346	44	4.76	4.76	0.785	33
8	370	46	9.76	9.76	0.553	35

Pada kedua kasus tersebut, model tetap menghasilkan ringkasan yang faktual terhadap dokumen sumber dengan menyertakan entitas numerik penting, namun memperoleh skor ROUGE rendah karena referensi manusia menggunakan strategi summarization yang lebih abstrak dan tidak

mempertahankan detail nilai Rupiah. Kondisi ini menunjukkan adanya *reference-summary mismatch* antara output model dan preferensi annotator.
Selain itu, korelasi tinggi antara jumlah token dokumen dan missing Rupiah menunjukkan bahwa semakin panjang dokumen, semakin banyak entitas numerik yang tidak tercantum pada ringkasan referensi. Temuan ini mengindikasikan bahwa evaluasi berbasis ROUGE pada dokumen finansial memiliki keterbatasan dalam menilai ketepatan informasi numerik secara kontekstual.
Secara umum, kasus kegagalan model tidak sepenuhnya disebabkan oleh kesalahan generasi, tetapi juga dipengaruhi oleh inkonsistensi strategi anotasi pada data referensi.

3.7 Analisis Kasus Terbaik dan Validasi Kualitas Output

Tabel 4 menunjukkan beberapa sampel dengan skor ROUGE-L tertinggi pada *Test Set*..

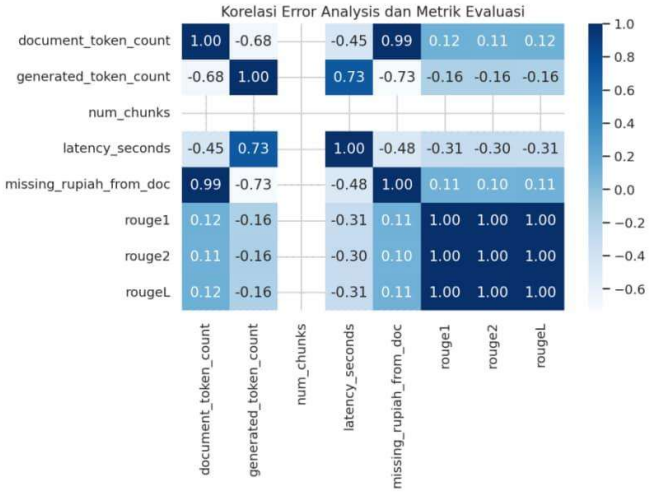
Table 4. Kasus Terbaik Berdasarkan Kualitas Kualitatif

Kabupaten	Doc Tokens	Gen Tokens	ROUGE-1	ROUGE-L	Latency (s)	Miss Rupiah
Pacitan	348	44	100	100	0.604	33
Magetan	328	45	100	100	0.582	31
Sumenep	312	45	100	100	0.542	28

Ketiga sampel menunjukkan kesesuaian penuh antara ringkasan hasil generasi dan referensi, dengan panjang ringkasan yang relatif konsisten pada rentang 44–45 token. Hasil ini menunjukkan bahwa model mampu mempelajari pola summarization yang berulang pada dokumen POK keuangan, terutama dalam mengekstraksi kode wilayah, nama kabupaten, dan entitas nilai Rupiah utama.
Temuan ini mengindikasikan bahwa *fine-tuning* BART cukup efektif pada domain dokumen yang memiliki struktur administratif semi-terstruktur dan pola linguistik yang konsisten.

3.8 Analisis Korelasi Antar-Variabel dan Implikasi Sistemik

Correlation heatmap pada Gambar 5 digunakan untuk mengamati hubungan antar variabel pada proses evaluasi dan *error analysis*. Korelasi tertinggi ditemukan antara document_token_count dan missing_rupiah_from_doc (r = 0,99), yang menunjukkan bahwa semakin panjang dokumen, semakin banyak entitas Rupiah yang tidak tercantum pada ringkasan referensi. Temuan ini mengindikasikan adanya keterbatasan pada konsistensi anotasi referensi, khususnya pada dokumen dengan kepadatan numerik tinggi.
Selain itu, terdapat korelasi positif antara generated_token_count dan latency_seconds (r = 0,73), yang menunjukkan bahwa panjang output berpengaruh langsung terhadap waktu inferensi. Sebaliknya, panjang dokumen input tidak menunjukkan hubungan linear terhadap latensi karena dokumen panjang diproses menggunakan mekanisme *chunking*.
Hasil lain menunjukkan bahwa variabel num_chunks memiliki korelasi yang relatif kecil terhadap skor ROUGE maupun latensi. Hal ini mengindikasikan bahwa mekanisme *chunking* yang diterapkan mampu menjaga stabilitas kualitas ringkasan tanpa memberikan dampak signifikan terhadap performa inferensi model.



Gambar 5. *correlation heatmap* antara Variabel *error analysis* dan Metrik Evaluasi

3.9 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, ukuran *Test Set* yang relatif kecil menyebabkan variansi skor evaluasi masih cukup tinggi. Kedua, terdapat inkonsistensi strategi anotasi pada ringkasan referensi, khususnya dalam penanganan entitas numerik, sehingga menciptakan target summarization yang ambigu bagi model.

Ketiga, evaluasi penelitian masih terbatas pada metrik ROUGE tanpa *human evaluation* maupun evaluasi faktual khusus untuk entitas numerik. Keempat, model hanya diuji pada dokumen POK keuangan di Provinsi Jawa Timur sehingga kemampuan generalisasi pada dokumen dari wilayah lain belum dapat dipastikan. Selain itu, pengujian latensi hanya dilakukan pada satu konfigurasi perangkat keras sehingga performa inferensi dapat berbeda pada lingkungan komputasi lainnya.

Penelitian ini menunjukkan bahwa model BART dapat diadaptasi untuk melakukan summarization pada dokumen POK keuangan berbahasa Indonesia yang memiliki struktur semi-terstruktur dan kepadatan entitas numerik tinggi. Dari sisi praktis, sistem memiliki potensi untuk diintegrasikan pada pengelolaan dokumen keuangan pemerintah daerah karena mampu menghasilkan ringkasan secara otomatis dengan latensi inferensi yang relatif rendah.

Selain itu, hasil *error analysis* menunjukkan bahwa kualitas anotasi referensi memiliki pengaruh besar terhadap hasil evaluasi model. Temuan ini mengindikasikan bahwa peningkatan kualitas dataset dan konsistensi anotasi dapat memberikan dampak signifikan terhadap performa summarization.

Pengembangan selanjutnya dapat difokuskan pada perluasan dataset, penggunaan model multilingual, penerapan evaluasi berbasis human assessment, serta pengembangan metrik evaluasi yang lebih sensitif terhadap ketepatan entitas numerik pada dokumen finansial.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil merancang sistem peringkasan otomatis dokumen POK keuangan berbasis BART menggunakan pendekatan *abstractive summarization*. Hasil eksperimen menunjukkan bahwa model mampu menghasilkan ringkasan yang konsisten terhadap struktur administratif dan entitas finansial dokumen POK, dengan performa inferensi yang masih layak untuk skenario semi-real-time.

Temuan penelitian menunjukkan bahwa performa model dipengaruhi oleh karakteristik dokumen POK yang semi-terstruktur dan repetitif. Selain itu, *error analysis* mengungkap bahwa beberapa kasus dengan skor ROUGE rendah tetap menghasilkan ringkasan yang faktual, sehingga evaluasi

berbasis ROUGE pada domain finansial memiliki keterbatasan dalam menangkap ketepatan informasi numerik secara kontekstual. Penelitian ini juga menunjukkan bahwa mekanisme *chunking mampu* menangani dokumen panjang tanpa menurunkan kualitas ringkasan secara signifikan.

Penelitian selanjutnya dapat difokuskan pada perluasan dataset, peningkatan konsistensi anotasi referensi, serta penerapan *human evaluation* dan evaluasi *factual consistency* untuk melengkapi keterbatasan ROUGE. Selain itu, eksperimen menggunakan model multilingual dan implementasi sistem pada lingkungan pengelolaan dokumen keuangan nyata juga perlu dilakukan untuk menguji kemampuan generalisasi dan performa operasional sistem.

5. DAFTAR RUJUKAN

[1] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017.

[2] S. Allahyari et al., "Text summarization techniques: a brief survey," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.

[3] K. El-Kassas, C. Salama, A. Rafea, and H. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Systems with Applications*, vol. 165, pp. 1–33, 2021.

[4] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," *in Proc. EMNLP-IJCNLP, Hong Kong, China*, 2019, pp. 3730–3740.

[5] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization," in *Proc. ICML, Virtual*, 2020, pp. 11328–11339.

[6] M. Lewis et al., "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proc. ACL, Online*, 2020, pp. 7871–7880.

[7] A. Ramesh and S. Sanampudi, "An automatic text summarization approach based on BART and transformer models," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 1–10, 2022.

[8] A. M. Elrefaie and A. M. Ahmed, "Deep transformer language models for Arabic text summarization," *IEEE Access*, vol. 10, pp. 1–15, 2022.

[9] M. Moradi, A. Dashti, and M. Samwald, "Abstractive text summarization using transformer models," *Information Processing & Management*, vol. 58, no. 3, pp. 1–17, 2021.

[10] M. Fajri, R. Hidayat, and A. Nugroho, "Abstractive English document summarization using BART model with chunk method," in *Proc. International Conference on Informatics and Computing*, 2023, pp. 1–6.

[11] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," in *Proc. EMNLP, Lisbon, Portugal*, 2015, pp. 379–389.

[12] F. Kryściński, B. McCann, C. Xiong, and R. Socher, "Evaluating the factual consistency of abstractive text summarization," in *Proc. EMNLP-IJCNLP, Hong Kong, China*, 2019, pp. 933–939.

[13] R. Koto, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP," in *Proc. COLING, Barcelona, Spain*, 2020, pp. 757–770.

[14] R. Koto, J. H. Lau, and T. Baldwin, "Indonesian abstractive summarization with pre-trained transformer models," in *Proc. AACL-IJCNLP, Online*, 2020, pp. 1–10.

[15] K. E. Dewi and N. I. Widiastuti, "Automatic summarization of Indonesian texts using a hybrid approach," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 15, no. 1, pp. 37–43, 2022.