PENGEMBANGAN MODEL TEXT SUMMARIZATION BERBASIS TEXT-TO-TEXT TRANSFER TRANSFORMER (T5) UNTUK JURNAL ILMIAH

Raihan Maulana, Said Iskandar Al Idrus, Hermawan Syahputra, Kana Saputra S, Debi Yandra Niska, Muhammad Anggie Januarsyah Daulay
Ilmu Komputer, Universitas Negeri Medan
Jalan William Iskandar Pasar V Deli Serdang, Indonesia
raihanmaulana@mhs.unimed.ac.id

ABSTRAK

Pesatnya pertumbuhan publikasi ilmiah meningkatkan kebutuhan akan sistem peringkasan otomatis untuk membantu pembaca memahami inti informasi secara efisien. Namun, kompleksitas struktur teks jurnal, khususnya yang mengandung anotasi numerik dan ekspresi matematis, menjadi tantangan dalam menghasilkan ringkasan yang akurat. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi model abstractive text summarization berbasis T5-small pada jurnal ilmiah berbahasa Indonesia. Metode yang digunakan meliputi pra-pemrosesan teks, tokenisasi, pembatasan panjang token, serta fine-tuning model dengan tiga skema pembagian data, yaitu 70:30, 80:20, dan 90:10. Evaluasi dilakukan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L untuk mengukur kesamaan antara ringkasan hasil model dan abstrak referensi. Hasil penelitian menunjukkan bahwa variasi proporsi data latih mempengaruhi performa model secara tidak linear. Skema 70:30 menghasilkan nilai ROUGE-L tertinggi pada dataset full teks, sedangkan skema 90:10 memberikan performa terbaik pada dataset dengan anotasi numerik dan matematis. Selain itu, keberadaan ekspresi matematis cenderung menurunkan kualitas ringkasan. Secara keseluruhan, model T5-small mampu menghasilkan ringkasan dengan tingkat kesamaan leksikal yang baik, namun masih dipengaruhi oleh karakteristik dan kompleksitas dataset.

Kata kunci : *text summarization, T5-small, abstractive summarization, jurnal ilmiah, ROUGE,*

1. PENDAHULUAN

Perkembangan publikasi ilmiah mengalami peningkatan yang signifikan baik secara nasional maupun internasional, yang ditandai dengan tingginya jumlah artikel yang diterbitkan setiap tahunnya. Berdasarkan data Scimago Journal & Country Rank, Indonesia menempati peringkat ke-38 dunia dengan total 376.908 publikasi dalam periode 1996–2023 [1].

Peningkatan ini memberikan dampak positif terhadap perkembangan ilmu pengetahuan, namun di sisi lain menimbulkan tantangan bagi akademisi dan peneliti dalam memahami literatur ilmiah secara efisien. Kompleksitas struktur teks jurnal serta penggunaan bahasa akademik yang padat dan teknis menyebabkan proses pemahaman memerlukan waktu dan usaha yang besar.

Permasalahan utama yang muncul adalah ketidakefisienan proses peringkasan manual serta keterbatasan pendekatan otomatis yang ada. Peringkasan manual memerlukan waktu yang lama, terutama ketika harus menganalisis banyak dokumen sekaligus. Sementara itu, sebagian besar penelitian sebelumnya dalam text summarization masih berfokus pada data berita atau teks umum serta menggunakan pendekatan ekstraktif, yang sering menghasilkan ringkasan kurang koheren. Selain itu, evaluasi model umumnya hanya menggunakan metrik ROUGE tanpa mempertimbangkan aspek keterbacaan dan koherensi. Hal ini menunjukkan adanya kesenjangan penelitian (research gap), khususnya dalam pengembangan model peringkasan abstraktif untuk teks jurnal ilmiah yang memiliki struktur kompleks dan karakteristik bahasa yang lebih formal.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan model text summarization berbasis pendekatan abstraktif menggunakan arsitektur Text-to-Text Transfer Transformer (T5) yang di-fine-tuning pada dataset jurnal ilmiah berbahasa Indonesia. Model ini dirancang untuk menghasilkan ringkasan yang lebih koheren, informatif, dan sesuai dengan karakteristik teks akademik. Rencana penyelesaian masalah dilakukan melalui tahapan pra-pemrosesan teks, tokenisasi, pembatasan panjang token, serta fine-tuning model dengan beberapa skema pembagian data. Evaluasi model tidak hanya menggunakan metrik ROUGE, tetapi juga mempertimbangkan kualitas ringkasan dari aspek keterbacaan dan kekoherenan. Dengan pendekatan tersebut, penelitian ini diharapkan dapat meningkatkan efisiensi pemahaman literatur ilmiah serta memberikan kontribusi dalam pengembangan metode peringkasan teks untuk dokumen akademik yang kompleks.

2. TINJAUAN PUSTAKA

2.1. Text Summarization

Text Summarization adalah proses yang melibatkan pemadatan teks dalam jumlah besar dan menjadi versi yang lebih pendek dengan tetap mempertahankan informasi dari makna yang penting. Tugas ini menjadi semakin penting karena pertumbuhan eksponensial konten tekstual di internet dan di berbagai arsip, seperti artikel berita, makalah ilmiah, dan dokumen hukum [2].

2.2. Abstractive Summarization

Abstractive Summarization adalah teknik dalam pemrosesan bahasa alami (NLP) yang bertujuan untuk menghasilkan ringkasan dari teks sumber dengan cara menulis ulang informasi penting dalam bentuk yang lebih ringkas dan mudah dibaca, sering kali menggunakan kosakata yang tidak ada dalam teks asli [3].

Tantangan yang dihadapi dalam melakukan Abstractive Summarization ialah menjaga konsistensi faktual, menangani struktur bahasa yang beragam dan 8 memastikan pentingnya infromasi. Meskipun matriks evaluasi seperti ROUGE umumnya sering di gunaka, tetap saja evaluasi manusia menjadi standar yang lebih bagus untuk menilai kualitas ringkasan [4].

Abstractive Summarization telah mengalami kemajuan yang signifikan dengan pengembangan model Deep Learning, terutama yang didasarkan pada arsitektur Sequence-to-Sequence dan model Transformer [5].

Model-model ini mampu memahami dan menghasilkan teks yang mirip manusia, sehingga cocok untuk tugas-tugas Abstractive Summarization [6].

2.3. Transformer

Transformer adalah arsitektur model deep learning yang telah diadopsi di berbagai bidang kecerdasan buatan seperti Natural Language Processing, Computer Vision, dan Speech Processing. Arsitektur Transformer terdiri dari encoder dan decoder, masing-masing terdiri dari tumpukan blok identik yang menggabungkan modul self-attention multi-head dan feed-forward network.

Self-attention adalah mekanisme kunci dalam arsitektur Transformer yang memungkinkan model untuk mempertimbangkan hubungan antara setiap pasangan elemen dalam input sequence. Mekanisme ini menghitung bobot perhatian untuk setiap elemen berdasarkan kesamaan dengan elemen lainnya, memungkinkan model untuk menangkap dependensi jangka panjang dalam data [7].

2.4. Text-to-Text Transfer Transformer

T5 (*Text-to-Text Transfer Transformer*) merupakan model berbasis transformer yang dirancang untuk mendukung pembelajaran multitugas dalam domain *Natural Language Processing* (NLP) [10].

Model ini menggunakan arsitektur *encoder-decoder* dengan mekanisme *self-attention* untuk menangani input berukuran variabel tanpa bergantung pada RNN atau CNN [8]. Setiap encoder terdiri dari lapisan *self-attention* dan *feed-forward*, sedangkan decoder memiliki mekanisme tambahan untuk memperhatikan output encoder.

Dibandingkan dengan RNN, T5 lebih efisien karena mendukung komputasi paralel serta mampu menangkap ketergantungan jarak jauh antar token yang penting dalam tugas peringkasan. T5 juga memiliki beberapa varian ukuran, yaitu T5-Small (60 juta parameter), T5-Base (220 juta), T5-Large (770 juta), T5-3B (3 miliar), dan T5-11B (11 miliar parameter) [9].

Dalam implementasinya, T5 menggunakan dua tahap utama, yaitu *pre-training* untuk membangun pengetahuan umum dan *fine-tuning* untuk menyesuaikan model pada tugas spesifik [10].

2.5. ROUGE

ROUGE (Recall-Oriented Understudy for Gisting Evaluation) adalah metrik yang digunakan untuk mengevaluasi kualitas ringkasan teks dengan membandingkannya dengan ringkasan emas (gold summary) yang dihasilkan oleh manusia. Metrik ini mengukur tingkat kesamaan berdasarkan unit yang tumpang tindih, seperti pasangan kata, n-gram, dan urutan kata, antara ringkasan yang dihasilkan oleh model dan ringkasan referensi [11].

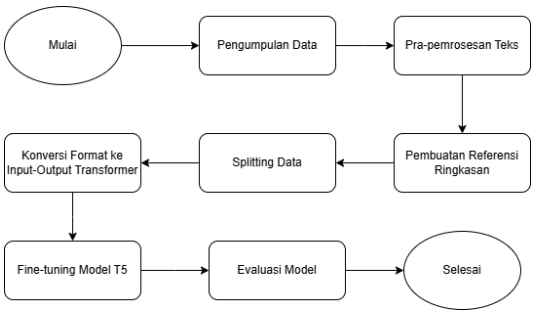
ROUGE memiliki beberapa varian, di antaranya ROUGE-N dan ROUGE L. ROUGE-N mengukur jumlah n-gram yang tumpang tindih antara ringkasan model dan referensi. ROUGE-L menggunakan Longest Common Subsequence 23 (LCS) untuk mengevaluasi kesamaan berdasarkan urutan kata terpanjang yang muncul secara berurutan dalam kedua ringkasan [11].

2.6. Penelitian Terdahulu

Penelitian terkait text summarization berbasis Transformer, khususnya model T5, telah menunjukkan perkembangan yang signifikan dalam beberapa tahun terakhir. Andika Bahari dan Kania Evita Dewi mengembangkan model abstractive summarization berbasis Transformer dengan membandingkan dua varian T5, yaitu T5-default dan T5-id. Hasil penelitian menunjukkan bahwa penggunaan kosakata khusus bahasa Indonesia (T5-id) mampu meningkatkan skor ROUGE meskipun menghasilkan validation loss yang lebih tinggi. Studi ini menegaskan bahwa pemilihan vocab dan strategi fine-tuning berpengaruh terhadap kualitas ringkasan yang dihasilkan[12].

Meskipun kedua penelitian tersebut menunjukkan efektivitas model T5 dalam peringkasan teks berbahasa Indonesia, keduanya masih berfokus pada data berita atau teks umum. Selain itu, evaluasi yang dilakukan umumnya hanya menggunakan metrik ROUGE tanpa mempertimbangkan aspek keterbacaan dan koherensi secara langsung. Oleh karena itu, penelitian ini mengembangkan model abstractive summarization berbasis T5 yang di-fine-tuning pada dataset jurnal ilmiah, termasuk teks yang mengandung anotasi numerik dan ekspresi matematis, serta melibatkan evaluasi tambahan melalui validasi ahli untuk menilai kualitas ringkasan secara lebih komprehensif.

3. METODE PENELITIAN



Gambar 1. Diagram Alur Proses Pengembangan dan Evaluasi Model Text Summarization Berbasis T5

3.1. Pengumpulan Data

Pada tahap awal penelitian, data dikumpulkan dari jurnal ilmiah *Open Access* yang dapat diakses secara legal untuk keperluan penelitian. Proses pengumpulan dilakukan secara manual dengan menyeleksi artikel berdasarkan kriteria tertentu, yaitu jurnal dengan teks penuh dari berbagai bidang ilmu tanpa mengandung formula matematis atau elemen visual kompleks seperti grafik dan tabel yang dapat mengganggu konteks naratif. Bagian yang digunakan sebagai data adalah seluruh isi jurnal, kecuali abstrak dan kesimpulan, untuk menghindari penggunaan ringkasan yang telah tersedia serta meminimalkan bias dalam proses pelatihan model.

3.2. Pra-pemrosesan Teks

Pada tahap ini, data teks yang telah dikumpulkan dibersihkan dan dipersiapkan sebelum digunakan dalam pelatihan model. Proses pra-pemrosesan meliputi penghapusan karakter atau simbol yang tidak relevan, seperti tanda baca tertentu dan elemen non-teks. Tahapan ini bertujuan untuk meningkatkan kualitas data masukan sehingga model dapat memahami struktur teks dengan lebih baik dan menghasilkan ringkasan yang lebih akurat.

3.3. Pembuatan Referensi Ringkasan

Setelah melalui tahap pra-pemrosesan, teks jurnal digunakan untuk menghasilkan ringkasan yang berfungsi sebagai referensi dalam proses *fine-tuning* model. Ringkasan ini digunakan sebagai target output sehingga model dapat mempelajari pola peringkasan dan menghasilkan keluaran yang mendekati kualitas ringkasan yang diharapkan.

3.4. Splitting Data

Dataset kemudian dibagi menjadi data pelatihan (*training*) dan data pengujian (*testing*) menggunakan tiga skenario rasio, yaitu 90:10, 80:20, dan 70:30. Variasi rasio ini digunakan untuk menganalisis pengaruh proporsi data terhadap performa model. Data pelatihan digunakan dalam proses *fine-tuning*, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya, sehingga dapat mengurangi risiko *overfitting*.

3.5. Konversi Format ke Input-Output Transformer

Pada tahap ini, data dikonversi ke dalam format yang sesuai dengan arsitektur T5 yang berbasis *text-to-text*. Input berupa teks jurnal hasil pra-pemrosesan, sedangkan output berupa ringkasan referensi. Proses konversi mencakup tokenisasi menggunakan tokenizer T5 untuk mengubah teks menjadi token ID, yang kemudian dipetakan ke dalam vektor embedding berdimensi tetap. Selain itu, dilakukan penyesuaian teknis seperti penambahan *special tokens* dan *padding* agar struktur data sesuai dengan kebutuhan model.

3.6. Fine-Tuning Model

Tahap *fine-tuning* bertujuan untuk menyesuaikan model T5 agar mampu memahami karakteristik teks jurnal ilmiah berbahasa Indonesia. Proses ini dilakukan dengan melatih model menggunakan pasangan input berupa teks jurnal dan output berupa ringkasan referensi. Selama pelatihan, model mempelajari pola-pola dalam data dan menyesuaikan parameter internalnya untuk menghasilkan ringkasan yang lebih relevan. Parameter seperti *learning rate*, *batch size*, dan jumlah *epoch* diatur untuk memperoleh performa optimal serta menghindari *overfitting*.

3.7. Evaluasi Model

Evaluasi dilakukan untuk mengukur kinerja model dalam menghasilkan ringkasan dengan membandingkan output model terhadap ringkasan referensi menggunakan metrik ROUGE. Metrik yang digunakan meliputi ROUGE-1, ROUGE-2, dan ROUGE-L, yang mengukur kesamaan berdasarkan n-gram dan urutan kata. Penilaian dilakukan melalui *precision*, *recall*, dan *F1-score* untuk mengevaluasi kemampuan model dalam menangkap informasi penting dari teks. Hasil evaluasi ini digunakan untuk menilai efektivitas model setelah proses *fine-tuning*.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Data penelitian terdiri atas dua jenis, yaitu jurnal ilmiah dengan teks lengkap serta jurnal yang mengandung angka dan anotasi matematis. Seluruh data diperoleh dari laman Garuda (Garba Rujukan Digital) melalui proses pengunduhan manual terhadap setiap dokumen. Jurnal yang dikumpulkan berasal dari berbagai bidang ilmu, antara lain *Social Sciences*, *Education*, *Law*, *Crime, Criminology & Criminal Justice*, *Humanities*, *Library & Information Science*, *Language, Linguistic, Communication & Media*, *Nursing*, *Arts*, *Economics, Econometrics & Finance*, *Computer Science*, *Public Health*, dan *Mathematics*, dengan total sebanyak 759 dokumen. Data kemudian dikonversi dari format PDF ke dalam bentuk teks (.txt) dengan mengekstraksi bagian judul, subjudul, dan isi, sementara bagian abstrak, kesimpulan, dan daftar

pustaka tidak disertakan untuk menghindari redundansi dan potensi bias terhadap ringkasan.

Selanjutnya, data disusun dalam format CSV dengan struktur kolom judul, bagian, dan isi_bagian untuk memudahkan proses pengolahan. Untuk menyesuaikan dengan kapasitas model, teks pada kolom isi_bagian dipecah menjadi segmen dengan panjang maksimum 512 token sehingga menghasilkan data yang lebih terstruktur dan siap digunakan dalam proses pelatihan. Selain itu, setiap segmen teks dilengkapi dengan ringkasan sebagai label yang disusun dengan mempertimbangkan struktur dan gaya penulisan ilmiah. Dengan demikian, dataset yang dihasilkan tidak hanya memuat pasangan teks dan ringkasan, tetapi juga mendukung model dalam mempelajari hubungan antara teks panjang dan bentuk ringkasnya secara sistematis.

4.2. Splitting Data

Data penelitian dikelompokkan ke dalam dua jenis, yaitu data *full text* dan data gabungan yang terdiri atas *full text* serta data yang memuat angka dan anotasi matematika. Pemisahan ini bertujuan untuk mengevaluasi performa model pada karakteristik data yang berbeda, yakni teks naratif murni dan teks dengan elemen numerik. Masing-masing dataset kemudian dibagi menggunakan tiga skenario rasio (70:30, 80:20, dan 90:10) untuk data latih, validasi, dan uji dengan nilai *seed* 42 guna memastikan konsistensi dan reprodusibilitas eksperimen.

Tabel 1. Pembagian Data Latih, Validasi, dan Uji (Full Teks)

No	Rasio	Data Latih	Data Validasi	Data Uji
1	70:30	2148	460	461
2	80:20	2455	307	307
3	90:10	2761	154	154

Pada dataset *full text* dengan total 3.069 entri, distribusi data menunjukkan bahwa semakin besar proporsi data latih, jumlah data uji dan validasi semakin kecil. Variasi rasio ini digunakan untuk mengamati pengaruh jumlah data latih terhadap performa model, khususnya dalam menangkap pola pada teks ilmiah yang bersifat naratif.

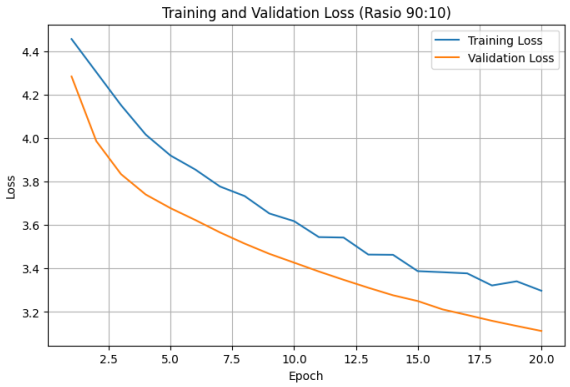
Tabel 2. Pembagian Data Latih, Validasi, dan Uji (Gabungan Teks dan Anotasi Matematika)

No	Rasio	Data Latih	Data Validasi	Data Uji
1	70:30	3096	662	664
2	80:20	3537	442	443
3	90:10	3979	221	222

4.3. Fine-Tuning Model

Pada tahap fine-tuning, model Transformer dilatih ulang menggunakan dataset penelitian untuk menyesuaikan kemampuannya dalam peringkasan teks ilmiah, dengan memanfaatkan DataCollatorForSeq2Seq untuk mengelola batching, padding, dan penyelarasan input-output secara dinamis. Konfigurasi pelatihan diatur menggunakan Seq2SeqTrainingArguments dengan learning rate sebesar 3e-5, batch size kecil yang dikombinasikan dengan gradient accumulation, serta evaluasi dan penyimpanan model pada setiap epoch. Pelatihan dilakukan selama 20 epoch dengan penerapan teknik regularisasi seperti weight decay, label smoothing, gradient clipping, dan learning rate warm-up guna menjaga stabilitas serta meningkatkan generalisasi model. Model terbaik dipilih berdasarkan nilai validation loss terendah, sementara performa dievaluasi menggunakan metrik ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L) serta generated length untuk menganalisis karakteristik ringkasan yang dihasilkan.

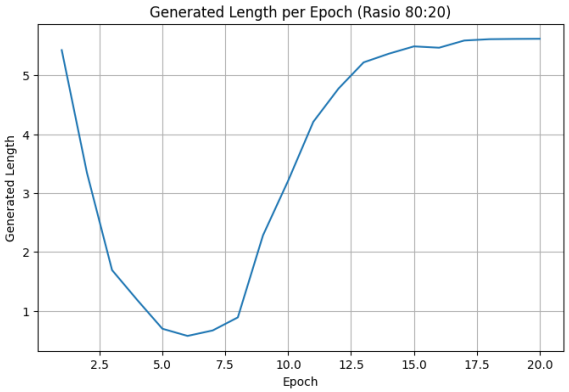
Hasil fine-tuning pada dataset full text dengan tiga rasio pembagian data (70:30, 80:20, dan 90:10) menunjukkan tren pembelajaran yang stabil, ditandai dengan penurunan konsisten pada Training Loss dan Validation Loss di seluruh epoch tanpa indikasi overfitting. Pada rasio 70:30, Validation Loss menurun hingga 3,2014, diikuti oleh 80:20 sebesar 3,1372, dan nilai terendah pada 90:10 sebesar 3,1123, yang menunjukkan bahwa peningkatan proporsi data latih berkontribusi terhadap peningkatan kemampuan generalisasi model. Dari sisi kualitas ringkasan, metrik ROUGE pada ketiga skenario menunjukkan pola serupa, yaitu penurunan pada fase awal pelatihan akibat proses adaptasi model, kemudian meningkat secara bertahap hingga akhir epoch. Rasio 90:10 menghasilkan performa terbaik dengan ROUGE-1 sebesar 0,0906, ROUGE-2 sebesar 0,0459, dan ROUGE-L sebesar 0,0822, lebih tinggi dibandingkan rasio 70:30 dan 80:20 yang menunjukkan peningkatan secara moderat. Sementara itu, generated length relatif stabil pada seluruh konfigurasi di akhir pelatihan, sehingga peningkatan performa tidak dipengaruhi oleh panjang ringkasan, melainkan oleh kualitas representasi yang dipelajari model. Secara keseluruhan, peningkatan proporsi data latih berkorelasi dengan peningkatan performa, meskipun interpretasi pada rasio 90:10 tetap perlu dilakukan secara hati-hati karena ukuran data validasi yang lebih kecil.



Gambar 2. Grafik training loss dan validation loss pada rasio 90:10 Dataset Full Teks

Grafik pada Gambar 1 menunjukkan penurunan loss yang konsisten tanpa fluktuasi signifikan, yang mengindikasikan proses pelatihan berlangsung stabil dan konvergen.

Hasil fine-tuning pada dataset gabungan yang mengandung angka dan anotasi matematika menunjukkan dinamika pelatihan yang lebih fluktuatif pada fase awal, namun tetap mencapai konvergensi yang stabil pada akhir pelatihan. Seluruh rasio pembagian data (70:30, 80:20, dan 90:10) memperlihatkan penurunan konsisten pada training loss dan validation loss tanpa indikasi overfitting, dengan nilai validation loss akhir masing-masing sebesar 3,1003, 3,0574, dan 3,0591, yang menunjukkan bahwa rasio 80:20 memberikan performa validasi terbaik meskipun selisihnya relatif kecil. Dari sisi kualitas ringkasan, metrik ROUGE pada ketiga konfigurasi menunjukkan pola serupa, yaitu penurunan tajam pada fase awal akibat adaptasi terhadap kompleksitas data, kemudian meningkat secara bertahap hingga akhir pelatihan, dengan performa akhir yang relatif kompetitif tanpa adanya dominasi konsisten dari satu rasio tertentu. Sementara itu, generated length pada seluruh konfigurasi cenderung stabil pada kisaran 5,6 token, sehingga variasi performa lebih dipengaruhi oleh kualitas representasi model dibandingkan panjang ringkasan. Secara keseluruhan, peningkatan proporsi data latih tidak selalu menghasilkan peningkatan performa yang linear pada dataset gabungan, di mana rasio 80:20 menunjukkan keseimbangan terbaik antara stabilitas pelatihan dan performa validasi, sedangkan rasio 90:10 hanya memberikan peningkatan marginal pada sebagian metrik. Gambar X menunjukkan kurva training loss dan validation loss pada rasio 80:20 sebagai representasi proses konvergensi model.



Gambar 3. Grafik training loss dan validation loss pada rasio 80:20 Dataset Full Teks dan Anotasi Matematika

Kurva pada Gambar X menunjukkan penurunan training loss dan validation loss yang konsisten tanpa fluktuasi signifikan, yang mengindikasikan proses konvergensi model berjalan stabil. Hal ini memperkuat bahwa rasio 80:20 memberikan keseimbangan yang baik antara pembelajaran dan kemampuan generalisasi pada dataset gabungan.

4.4. Evaluasi Model

Evaluasi model dilakukan untuk mengukur kualitas ringkasan yang dihasilkan menggunakan metrik ROUGE, yang mencakup ROUGE-1, ROUGE-2, dan ROUGE-L. Pengujian dilakukan pada data uji dengan tiga skenario pembagian data (70:30, 80:20, dan 90:10) untuk membandingkan performa model secara konsisten pada kondisi yang sama.

Tabel 3. Hasil Evaluasi Model pada Dataset Full Teks

No	Model Rasio	ROUGE-1	ROUGE-2	ROUGE-L
1	70:30	0.424946	0.168312	0.254098
2	80:20	0.424328	0.166180	0.253271
3	90:10	0.423378	0.166417	0.252686

Berdasarkan hasil evaluasi pada dataset full text, performa model pada ketiga rasio pembagian data menunjukkan hasil yang relatif berdekatan. Model dengan rasio 70:30 memperoleh nilai tertinggi pada seluruh metrik ROUGE, meskipun selisihnya terhadap rasio 80:20 dan 90:10 tidak signifikan. Hal ini menunjukkan bahwa peningkatan proporsi data latih tidak secara langsung menghasilkan peningkatan performa pada tahap pengujian. Secara umum, konfigurasi 70:30 memberikan keseimbangan yang lebih baik antara proses pembelajaran dan kemampuan generalisasi, khususnya dalam mempertahankan kesesuaian unigram, bigram, serta struktur informasi dalam ringkasan.

Tabel 4. Hasil Evaluasi Model pada Dataset Full Teks dan Anotasi Matematika

No	Model Rasio	ROUGE-1	ROUGE-2	ROUGE-L
1	70:30	0.404021	0.161839	0.240973
2	80:20	0.404794	0.162963	0.240983
3	90:10	0.401421	0.163155	0.241688

Pada dataset gabungan yang mengandung angka dan anotasi matematika, performa model juga menunjukkan nilai yang relatif berdekatan antar rasio. Tidak terdapat satu konfigurasi yang secara konsisten unggul pada seluruh metrik, di mana rasio 80:20 menghasilkan nilai tertinggi pada ROUGE-1, sedangkan rasio 90:10 sedikit unggul pada ROUGE-2 dan ROUGE-L. Perbedaan nilai yang sangat tipis ini menunjukkan bahwa peningkatan proporsi data latih tidak memberikan dampak yang signifikan terhadap peningkatan kualitas ringkasan secara keseluruhan. Selain itu, variasi performa antar metrik mengindikasikan bahwa kompleksitas data, khususnya keberadaan elemen numerik dan anotasi matematis, mempengaruhi stabilitas model dalam menangkap informasi secara konsisten.

4.5. Validasi Ahli

Validasi kualitas ringkasan dilakukan oleh satu orang ahli bahasa menggunakan instrumen berbasis

skala Likert (1–5) yang mencakup lima aspek utama, yaitu relevansi dan kelengkapan informasi (30%), akurasi (30%), koherensi (15%), kepadatan (15%), serta keterbacaan (10%). Penilaian dilakukan terhadap 20 data uji, dengan nilai akhir dihitung menggunakan rata-rata berbobot dari seluruh aspek. Hasil evaluasi menunjukkan bahwa model memperoleh nilai rata-rata sebesar 4,0225 yang berada pada kategori baik, dengan standar deviasi sebesar 0,3665 yang mengindikasikan bahwa variasi penilaian relatif kecil dan performa model cenderung konsisten dalam menghasilkan ringkasan.

Secara distribusi, sebagian besar nilai berada pada rentang 3,5 hingga 4,6, dengan nilai tertinggi sebesar 4,6 dan terendah 3,3. Aspek akurasi dan relevansi umumnya memperoleh skor lebih tinggi, menunjukkan bahwa model mampu mempertahankan informasi penting dari teks sumber secara memadai. Namun demikian, pada beberapa data masih ditemukan nilai yang lebih rendah pada aspek koherensi dan kepadatan, yang mengindikasikan bahwa struktur penyampaian dan efisiensi informasi belum sepenuhnya optimal. Meskipun hasil validasi menunjukkan performa yang baik dan stabil, penggunaan satu validator dalam penelitian ini menjadikan evaluasi bersifat terbatas dan eksploratif, sehingga hasil yang diperoleh perlu diinterpretasikan secara hati-hati.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model *text summarization* berbasis T5-small berhasil di-*fine-tuning* menggunakan dataset full teks jurnal ilmiah berbahasa Indonesia dengan abstrak sebagai referensi melalui tahapan pra-pemrosesan, tokenisasi, pembatasan panjang token, serta format *text-to-text* sesuai arsitektur T5. Pengujian menggunakan tiga skema pembagian data (70:30, 80:20, dan 90:10) menunjukkan bahwa proporsi data latih mempengaruhi performa model, meskipun peningkatan performa tidak bersifat linear pada seluruh metrik evaluasi. Pada dataset full teks, nilai ROUGE-L tertinggi diperoleh pada skema 70:30, sedangkan pada dataset yang mengandung anotasi numerik dan matematis, skema 90:10 secara umum memberikan performa terbaik. Perbedaan ini menunjukkan bahwa karakteristik dan kompleksitas teks, khususnya keberadaan ekspresi matematis dan simbol numerik, mempengaruhi stabilitas pelatihan dan kualitas ringkasan yang dihasilkan. Secara keseluruhan, model T5-small mampu menghasilkan ringkasan dengan tingkat kesamaan leksikal yang cukup baik terhadap abstrak referensi, namun performanya masih dipengaruhi oleh distribusi data latih dan struktur teks ilmiah.

Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi model dengan kapasitas parameter yang lebih besar atau arsitektur transformer lainnya, menambahkan metode evaluasi berbasis penilaian manusia untuk mengukur aspek koherensi dan keterbacaan, melakukan analisis lebih mendalam terhadap pengaruh pembagian data terhadap metrik evaluasi, mengembangkan teknik pra-pemrosesan khusus untuk menangani ekspresi numerik dan matematis, serta memperluas variasi dataset guna meningkatkan kemampuan generalisasi model.

DAFTAR PUSTAKA

[1] Scimago. (n.d.). SJR — scimago Journal & Country Rank [Portal]. Retrieved March 9, 2025, from http://www.scimagojr.com/

[2] El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. Expert Systems with Applications, 165, 113679. https://doi.org/10.1016/j.eswa.2020.113679

[3] Kirmani, M., Kaur, G., & Mohd, M. (2024). Analysis of Abstractive and Extractive Summarization Methods. International Journal of Emerging Technologies in Learning (iJET), 19(01), 86–96. https://doi.org/10.3991/ijet.v19i01.46079

[4] Rennard, V., Shang, G., Hunter, J., & Vazirgiannis, M. (n.d.). Abstractive Meeting Summarization: A Survey. https://doi.org/10.1162/tacl

[5] 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

[6] Alomari, A., Idris, N., Sabri, A. Q. M., & Alsmadi, I. (2022). Deep reinforcement and transfer learning for abstractive text summarization: A review. In Computer Speech and Language (Vol. 71). Academic Press. https://doi.org/10.1016/j.csl.2021.101276

[7] Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. AI Open, 3, 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

[8] Mastropaolo, A., Scalabrino, S., Cooper, N., Palacio, D. N., Poshyvanyk, D., Oliveto, R., & Bavota, G. (2021). Studying the Usage of Text-To-Text Transfer Transformer to Support Code-Related Tasks. http://arxiv.org/abs/2102.02017

[9] Ay, B., Ertam, F., Fidan, G., & Aydin, G. (2023). Turkish abstractive text document summarization using text to text transfer transformer. Alexandria Engineering Journal, 68, 1–13. https://doi.org/10.1016/j.aej.2023.01.008

[10] Itsnaini, Q. A., Hayaty, M., Putra, A. D., & Jabari, N. A. M. (2023). Abstractive Text Summarization using Pre-Trained Language Model "Text-to-Text Transfer Transformer (T5)." ILKOM Jurnal Ilmiah, 15(1), 124–131. https://doi.org/10.33096/ilkom.v15i1.1532.124-131

[11] Ibrahim, A., Alfonse, M., & Aref, M. (2023). A SYSTEMATIC REVIEW ON TEXT SUMMARIZATION OF MEDICAL RESEARCH ARTICLES. International Journal of Intelligent Computing and Information Sciences, 23(2), 50–61. https://doi.org/10.21608/ijicis.2023.190004.1252

[12] Bahari, A., & Dewi, K. E. (2024). PERINGKASAN TEKS OTOMATIS ABSTRAKTIF MENGGUNAKAN TRANSFORMER PADA TEKS BAHASA INDONESIA. 13(1). https://doi.org/10.34010/komputa.v13i1.11197