BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian mengenai pengembangan model text summarization berbasis T5-small untuk peringkasan jurnal ilmiah berbahasa Indonesia, dapat disimpulkan bahwa model T5-small berhasil di-fine-tuning menggunakan dataset full teks jurnal ilmiah sebagai input dan abstrak jurnal sebagai referensi ringkasan melalui tahapan pra-pemrosesan, tokenisasi, pembatasan panjang token, serta format text-to-text sesuai arsitektur T5. Hasil pengujian dengan tiga skema pembagian data (70:30, 80:20, dan 90:10) menunjukkan bahwa proporsi data latih mempengaruhi performa model, namun peningkatan performa tidak sepenuhnya bersifat linear pada seluruh metrik evaluasi.

Pada dataset full teks, nilai ROUGE menunjukkan variasi antar skema pembagian data, di mana ROUGE-L tertinggi diperoleh pada skema 70:30. Sementara itu, pada dataset full teks yang mengandung anotasi numerik dan matematis, konfigurasi 90:10 secara umum menghasilkan nilai ROUGE tertinggi. Perbedaan pola performa antara kedua dataset menunjukkan bahwa karakteristik dan kompleksitas teks berpengaruh terhadap stabilitas hasil pelatihan. Keberadaan ekspresi matematis dan simbol numerik menambah tingkat kompleksitas sehingga mempengaruhi kualitas ringkasan yang dihasilkan model. Secara keseluruhan, model T5-small mampu menghasilkan ringkasan dengan tingkat kesamaan leksikal yang cukup baik terhadap abstrak referensi berdasarkan metrik ROUGE. Namun, hasil eksperimen juga menunjukkan bahwa performa model masih dipengaruhi oleh distribusi data latih serta karakteristik struktur teks ilmiah.

5.2 Saran

Berdasarkan hasil penelitian dan temuan pada tahap evaluasi, beberapa saran untuk pengembangan penelitian selanjutnya adalah:

1. Menggunakan model dengan kapasitas parameter yang lebih besar atau arsitektur transformer lainnya untuk menganalisis pengaruh kompleksitas model terhadap stabilitas performa pada berbagai skema pembagian data.
2. Menambahkan metode evaluasi selain ROUGE, seperti evaluasi berbasis penilaian manusia, untuk mengukur aspek koherensi, keterbacaan, dan kesesuaian makna yang tidak sepenuhnya tercermin dalam kesamaan leksikal.
3. Melakukan analisis lebih mendalam terhadap pengaruh pembagian data latih terhadap variasi nilai ROUGE, terutama pada metrik ROUGE-L yang menunjukkan pola tidak sepenuhnya linear.
4. Mengembangkan teknik pra-pemrosesan khusus untuk menangani ekspresi numerik dan matematis agar model dapat lebih efektif dalam memahami struktur teks ilmiah yang kompleks.
5. Memperluas jumlah dan variasi dataset dari berbagai bidang keilmuan guna meningkatkan kemampuan generalisasi model terhadap domain yang berbeda.