

## ABSTRAK

Rafii Muhammad Fadhil, 51421205

PERBANDINGAN KINERJA MODEL BERTSUMEXT DENGAN BASELINE LEAD-3 UNTUK PERINGKASAN TEKS OTOMATIS BAHASA INGGRIS DAN BAHASA INDONESIA DENGAN PYTHON

Skripsi, Fakultas Teknologi Industri, Jurusan Informatika, Universitas Gunadarma, 2025.

Kata Kunci : *Automatic Text Summarization*, BERTSumExt, Lead-3, IndoSum, CNN/DailyMail, ROUGE, Python.

(xii + 73 + lampiran)

Peningkatan jumlah informasi digital mendorong kebutuhan akan sistem *Automatic Text Summarization* (ATS) yang mampu menyajikan ringkasan secara ringkas dan informatif. Penelitian ini bertujuan untuk membandingkan kinerja dua pendekatan ekstraktif, yaitu baseline sederhana Lead-3 dan model modern BERTSumExt berbasis Transformer, dengan menggunakan Python dalam merangkum teks berbahasa Inggris dan Bahasa Indonesia. Dataset yang digunakan adalah CNN/DailyMail sebagai representasi teks semi-abstraktif berbahasa Inggris dan IndoSum sebagai dataset standar berbahasa Indonesia yang bersifat ekstraktif. Tahapan penelitian mencakup pra-pemrosesan teks, strategi pelabelan (*fuzzy matching* untuk CNN/DailyMail dan *exact matching* untuk IndoSum), implementasi metode, pelatihan terbatas, serta evaluasi menggunakan metrik ROUGE. Hasil eksperimen menunjukkan bahwa pada CNN/DailyMail, Lead-3 memperoleh skor ROUGE-1 sebesar 0,2921, ROUGE-2 sebesar 0,1117, dan ROUGE-L sebesar 0,1915, lebih tinggi dibandingkan BERTSumExt yang hanya mencapai ROUGE-1 sebesar 0,2388, ROUGE-2 sebesar 0,0569, dan ROUGE-L sebesar 0,1427. Pada IndoSum, perbedaan semakin jelas, dengan Lead-3 mencapai ROUGE-1 sebesar 0,6687, ROUGE-2 sebesar 0,6046, dan ROUGE-L sebesar 0,6515, sedangkan BERTSumExt tertinggal dengan ROUGE-1 sebesar 0,3327, ROUGE-2 sebesar 0,1822, dan ROUGE-L sebesar 0,2442. Temuan ini menegaskan bahwa efektivitas model sangat dipengaruhi oleh karakteristik dataset; metode sederhana seperti Lead-3 lebih efektif pada dataset ekstraktif, sementara BERTSumExt tetap menawarkan fleksibilitas semantik pada dataset semi-abstraktif meskipun belum melampaui baseline. Penelitian ini berkontribusi dengan menghadirkan studi komparatif lintas bahasa mengenai efektivitas metode ekstraktif sederhana dan modern dalam peringkasan teks otomatis.

Daftar Pustaka (2017-2025)