# ABSTRAK

**Raihan, NIM 4213250004 (2026). Pengembangan Model Text Summarization Berbasis Text-to-Text Transfer Transformer (T5) untuk Jurnal Ilmiah.**

Perkembangan publikasi ilmiah yang pesat menyebabkan meningkatnya kebutuhan terhadap sistem peringkasan otomatis untuk membantu pembaca memahami inti jurnal secara efisien. Penelitian ini bertujuan untuk mengimplementasikan dan mengevaluasi model abstractive text summarization berbasis T5-small pada jurnal ilmiah berbahasa Indonesia. Dataset yang digunakan terdiri dari dua jenis, yaitu full teks jurnal dan full teks yang mengandung anotasi numerik serta ekspresi matematis, dengan abstrak jurnal sebagai referensi ringkasan. Proses penelitian meliputi tahapan pra-pemrosesan teks, tokenisasi, pembatasan panjang token, serta fine-tuning model menggunakan 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 tingkat kesamaan antara ringkasan hasil model dan abstrak referensi. Hasil pengujian menunjukkan bahwa proporsi data latih mempengaruhi performa model, namun peningkatan nilai ROUGE tidak sepenuhnya bersifat linear pada seluruh metrik. Pada dataset full teks, nilai ROUGE-L tertinggi diperoleh pada skema 70:30, sedangkan pada dataset yang mengandung anotasi numerik dan matematis, konfigurasi 90:10 secara umum menghasilkan performa terbaik. Selain itu, performa pada dataset yang mengandung ekspresi matematis cenderung lebih rendah dibandingkan full teks biasa, menunjukkan bahwa kompleksitas struktur teks mempengaruhi kualitas ringkasan yang dihasilkan. Secara keseluruhan, model T5-small mampu menghasilkan ringkasan dengan tingkat kesamaan leksikal yang cukup baik terhadap abstrak referensi, meskipun performanya masih dipengaruhi oleh karakteristik dataset dan distribusi data latih.

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

# ABSTRACT

**Raihan, NIM 4213250004 (2026). Development of a Text Summarization Model Based on the Text-to-Text Transfer Transformer (T5) for Scientific Journals.**

The rapid growth of scientific publications has increased the need for automatic summarization systems to help readers efficiently understand the core content of academic articles. This study aims to implement and evaluate an abstractive text summarization model based on T5-small for Indonesian scientific journals. The dataset consists of two types: full-text journal articles and full-text articles containing numerical annotations and mathematical expressions, with journal abstracts used as reference summaries. The research process includes text preprocessing, tokenization, sequence length limitation, and model fine-tuning using three data split configurations: 70:30, 80:20, and 90:10. The model performance was evaluated using ROUGE-1, ROUGE-2, and ROUGE-L metrics to measure the lexical similarity between the generated summaries and the reference abstracts. The experimental results indicate that the proportion of training data affects model performance; however, the improvement in ROUGE scores is not entirely linear across all evaluation metrics. For the full-text dataset, the highest ROUGE-L score was obtained under the 70:30 split configuration. In contrast, for the dataset containing numerical and mathematical annotations, the 90:10 configuration generally produced the best performance. Furthermore, the model's performance on texts containing mathematical expressions was slightly lower than on plain full-text articles, suggesting that textual complexity influences summarization quality. Overall, the T5-small model is capable of generating summaries with a reasonably good level of lexical similarity to the reference abstracts, although its performance remains influenced by dataset characteristics and training data distribution.

**Keywords:** text summarization, T5-small, abstractive summarization, scientific journals, ROUGE.