

**PERINGKASAN TEKS OTOMATIS MENGGUNAKAN METODE
BIDIRECTIONAL ENCODER REPRESENTATIONS FROM
TRANSFORMERS (BERT) PADA ARTIKEL BERITA POLITIK
BERBAHASA INDONESIA**

Oleh

Elma Margaretha Br Sebayang, NIM 2015101018

Jurusan Teknik Informatika

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi performa model *BERT-SUM* dan *DISTILBERT*, dalam menghasilkan ringkasan yang relevan dari artikel berita Politik berbahasa Indonesia. Penelitian ini berfokus pada tantangan peringkasan konten berita yang panjang dan kompleks, khususnya dalam bahasa Indonesia, menggunakan *peringkasan ekstraktif otomatis*, dan mengukur kualitas ringkasan menggunakan metrik evaluasi *ROUGE*. Dalam era digital, jumlah berita Politik yang tersedia secara daring semakin meningkat, menyebabkan tantangan bagi pembaca dalam memilah informasi yang relevan dan memahami inti dari suatu berita. Penelitian ini menggunakan pendekatan peringkasan *ekstraktif* dengan model *BERT-SUM* dan *DISTILBERT* untuk menghasilkan ringkasan dari artikel berita Politik Indonesia. Model-model ini dilatih pada dataset berita Indonesia, dan performanya dievaluasi menggunakan metrik *ROUGE*. Studi ini membandingkan kualitas dan relevansi ringkasan yang dihasilkan dengan ringkasan yang dibuat oleh manusia sebagai *ground truth*. Hasil evaluasi menunjukkan bahwa *BERT-SUM* unggul dibandingkan *DISTILBERT* pada semua metrik *ROUGE* di dataset Politik, dengan nilai *F1-Score* yang lebih tinggi pada *ROUGE-1*, *ROUGE-2*, dan *ROUGE-L*. *BERT-SUM* lebih efektif untuk peringkasan otomatis *ekstraktif* berita Politik Indonesia, sedangkan *DISTILBERT* dapat dipertimbangkan untuk lingkungan yang terbatas sumber daya komputasi.

Kata Kunci: *BERT-SUM*, *DISTILBERT*, *Peringkasan Teks Otomatis*, *Berita Politik Indonesia*, *ROUGE*, *Peringkasan Ekstraktif*.

***AUTOMATIC TEXT SUMMARIZATION USING BIDIRECTIONAL ENCODER
REPRESENTATIONS FROM TRANSFORMERS (BERT) ON INDONESIAN
POLITICAL NEWS ARTICLES***

By

Elma Margaretha Br Sebayang, NIM 2015101018

Department of Informatics Engineering

ABSTRACT

This research aims to evaluate the performance of the BERT-SUM and DISTILBERT models in generating relevant summaries of Indonesian political news articles. The study focuses on the challenges of summarizing long and complex news content, particularly in the Indonesian language, using automatic extractive summarization and assessing summary quality using the ROUGE evaluation metric. In the digital era, the increasing availability of political news online presents a challenge for readers in filtering relevant information and understanding the core message of a news article. This research employs an extractive summarization approach using BERT-SUM and DISTILBERT models to generate summaries of Indonesian political news articles. These models are trained on Indonesian news datasets, and their performance is evaluated using ROUGE metrics. The study compares the quality and relevance of the generated summaries with human-written summaries as ground truth. The evaluation Results indicate that BERT-SUM outperforms DISTILBERT across all ROUGE metrics on the political news dataset, achieving higher F1-Scores in ROUGE-1, ROUGE-2, and ROUGE-L. BERT-SUM proves to be more effective for automatic extractive summarization of Indonesian political news, while DISTILBERT may be considered for resource-constrained environments.

Keywords: BERT-SUM, DISTILBERT , automatic Text Summarization, Indonesian political news, ROUGE, extractive summarization.