

ABSTRAK

Dokumen berita berbahasa Indonesia umumnya memiliki panjang teks yang besar, redundansi informasi, dan struktur kalimat yang kompleks sehingga meningkatkan beban komputasi pada model klasifikasi berbasis Transformer. Penelitian ini bertujuan menganalisis pengaruh peringkasan ekstraktif TextRank terhadap kualitas ringkasan, performa klasifikasi, dan efisiensi komputasi pada klasifikasi berita berbahasa Indonesia menggunakan IndoRoBERTa. Peringkasan menggunakan TextRank dengan rasio 30%, 40%, dan 50% dari teks asli, baik tanpa maupun dengan preprocessing. Kualitas ringkasan dievaluasi menggunakan ROUGE-1, ROUGE-2, dan ROUGE-L dengan pseudo-reference yang dihasilkan oleh QuillBot Premium. Performa klasifikasi diukur menggunakan accuracy, precision, recall, dan F1-score, sedangkan efisiensi komputasi dianalisis berdasarkan waktu inferensi. Hasil penelitian menunjukkan bahwa kombinasi preprocessing dan TextRank menghasilkan kualitas ringkasan terbaik pada rasio 40%, dengan ROUGE-1 sebesar 0,5134, ROUGE-2 sebesar 0,2919, dan ROUGE-L sebesar 0,3209. Pada klasifikasi, teks asli (baseline) menghasilkan accuracy 96,45% dan F1-score tertinggi 96,44%. Sementara itu, skenario peringkasan terbaik diperoleh pada preprocessing + TextRank 50% dengan accuracy 94,98% dan F1-score 94,99%, atau mengalami penurunan accuracy sebesar 1,47% dibandingkan baseline. Dari sisi efisiensi, TextRank mampu menurunkan waktu inferensi dari 26,14 detik menjadi 22,88 detik atau lebih cepat 12,47%. Hasil ini menunjukkan bahwa TextRank efektif sebagai metode reduksi dimensi untuk meningkatkan efisiensi komputasi dengan penurunan performa klasifikasi yang relatif kecil.

Kata kunci: Peringkasan Ekstraktif, TextRank, Reduksi Dimensi Teks, Klasifikasi Berita, RoBERTa, Efisiensi Komputasi.

ABSTRACT

Indonesian language news documents generally have large text lengths, information redundancy, and complex sentence structures, which increase the computational burden on Transformer-based classification models. This study aims to analyze the effect of TextRank extractive summarization on summary quality, classification performance, and computational efficiency in Indonesian language news classification using IndoRoBERTa. Summarization uses TextRank with ratios of 30%, 40%, and 50% of the original text, both without and with preprocessing. Summary quality is evaluated using ROUGE-1, ROUGE-2, and ROUGE-L with pseudo-references generated by QuillBot Premium. Classification performance is measured using accuracy, precision, recall, and F1-score, while computational efficiency is analyzed based on inference time. The results show that the combination of preprocessing and TextRank produces the best summary quality at a ratio of 40%, with ROUGE-1 at 0.5134, ROUGE-2 at 0.2919, and ROUGE-L at 0.3209. In classification, the original text (baseline) produced an accuracy of 96.45% and the highest F1-score of 96.44%. Meanwhile, the best summarization scenario was obtained with preprocessing + TextRank 50% with an accuracy of 94.98% and an F1-score of 94.99%, or a decrease in accuracy of 1.47% compared to the baseline. In terms of efficiency, TextRank was able to reduce inference time from 26.14 seconds to 22.88 seconds or 12.47% faster. These results indicate that TextRank is effective as a dimensionality reduction method to improve computational efficiency with a relatively small decrease in classification performance.

Keywords: *Extractive Summarization, TextRank, Text Dimension Reduction, News Classification, RoBERTa, Computational Efficiency.*