ABSTRAK

Peningkatan produksi informasi digital yang masif menyebabkan fenomena *information overload* yang membuat manusia kesulitan menyerap informasi secara efisien, sehingga dibutuhkan teknologi peringkasan teks otomatis yang andal. Penelitian ini bertujuan untuk menganalisis perbandingan kinerja dan efisiensi komputasi tiga metode pelatihan model bahasa, yaitu *Full Fine-Tuning* (FFT), *Low-Rank Adaptation* (LoRA), dan *Knowledge Distillation* dengan LoRA (KD-LoRA), pada model IndoT5 untuk tugas peringkasan berita abstraktif berbahasa Indonesia. Ketiga metode diterapkan pada dataset IndoSum dengan pembagian data 64% untuk pelatihan, 16% untuk validasi, dan 20% untuk pengujian, menggunakan panjang token masukan 512 dan token keluaran 150. Hasil evaluasi menunjukkan bahwa KD-LoRA menghasilkan performa terbaik pada seluruh metrik kualitas, dengan ROUGE-1 sebesar 65,04, BERT*Score* F1 sebesar 84,61, serta *Novelty* 1-*gram* sebesar 10,26%, mengungguli FFT dengan selisih ROUGE-1 sebesar 0,69 poin dan BERT*Score* F1 sebesar 0,20 poin, serta mengungguli LoRA dengan selisih ROUGE-1 sebesar 0,88 poin dan BERT*Score* F1 sebesar 0,26 poin. Dari sisi efisiensi, KD-LoRA hanya melatih 1,24% dari total parameter (berkurang 98,77% dibanding FFT dan 61,10% dibanding LoRA), dengan FLOPs inferensi sebesar 9,71G (sekitar 76,95% lebih rendah dibanding FFT dan LoRA), ukuran *checkpoint* hanya 13,60 MB (berkurang 98,56% dibanding FFT dan 93,77% dibanding LoRA), rata-rata waktu pelatihan per *epoch* lebih singkat sebesar 1,82 menit dibanding FFT (2,25 menit) dan LoRA (2,37 menit), *peak* VRAM lebih rendah sekitar 47,40% dibanding FFT dan 43,34% dibanding LoRA, serta waktu inferensi lebih cepat sekitar 22,18% dibanding FFT dan 23,22% dibanding LoRA, sehingga menjadi metode paling seimbang antara kualitas ringkasan dan efisiensi komputasi untuk tugas peringkasan teks abstraktif bahasa Indonesia.

Kata kunci : peringkasan teks abstraktif, IndoT5, *Full Fine-Tuning*, LoRA, KD-LoRA, IndoSum