

Perbandingan Performansi Distilbert Indonesia dengan IndoBert dalam Ringkasan Ekstraktif

Hansel Nathanael Pinondang¹, Dr.Ir. MOCH. Arif Bijaksana, M.Tech.², Siti Sa'adah, S.T., M.T.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

⁴Divisi Digital Service PT Telekomunikasi Indonesia

¹hanselnp@students.telkomuniversity.ac.id, ²arifbijaksana@telkomuniversity.ac.id,

³sitisaadah@telkomuniversity.ac.id

Abstrak

- Dengan semakin bertambah majunya teknologi permodelan NLP, aplikasi untuk menyelesaikan masalah pun bertambah. Dan dari teknologi struktur model tersebut, muncul model menggunakan struktur yang memiliki kapabilitas yang mirip, tapi menggunakan sumber daya komputasi yang lebih sedikit. Studi kasus jurnal ini berupa berita yang dikumpulkan menjadi dataset benchmark dalam Bahasa Indonesia untuk pelatihan dalam tugas meringkas secara ekstraktif. Dataset yang dipakai adalah liputan6 dan indosum. Kedua model Bert and distilBert digunakan untuk melakukan *downward task* tersebut untuk mengukur performansi. Tolak ukur yang dipakai adalah total waktu training, dan penggunaan sumber daya komputasi serta *training loss* dari *fine-tuning*. Hasil dari *fine-tuning* menunjukkan nilai akhir *training loss* dari model Bert dengan distilBert tidak jauh berbeda dengan distilBert memerlukan waktu lebih cepat dan sumber komputasi lebih sedikit.

Kata kunci : DistilBERT, IndoBERT, Ekstraktif, Ringkasan, Transformers, NLP

Abstract

- With the advancement of NLP modeling technology, applications for problem solving have increased. And from model structure technology, models that use structures that have similar capabilities, but use less computational resources, are emerging. The case study of this journal is news collected into a benchmark dataset in Bahasa Indonesia for training extractive summarization tasks. The datasets used are liputan6 and indosum. Both Bert and distilBert models are used to perform the derivation task to measure performance. The benchmarks used are total training time, computational resource utilization, and training loss from fine-tuning. The results of fine-tuning show that the final training loss values of Bert and distilBert models are not much different with distilBert requiring faster time and less computational resources.

Keywords: DistilBERT, IndoBERT, Extractive, Summarization, Transformers, NLP

1. Pendahuluan

Latar Belakang

Recurrent neural network (RNN) atau variasi dari RNN telah terbukti efisien untuk tugas klasifikasi dan generasi sekuensi [12] dalam *Natural language Processing* (NLP). RNN berbeda dengan struktur lapisan lain dalam jaringan hirarkis karena adanya propagasi horisontal di antara node-node dalam lapisan yang sama. Perambatan ini menghubungkan output RNN dengan urutan input dan informasi sebelumnya. Dengan beberapa preprocessing, RNN mampu memodelkan urutan dengan panjang yang bervariasi. Tetapi, ketika skala ketergantungan jangka panjang yang harus dipelajari cukup besar, RNN sulit untuk dilatih dengan benar. RNN konvensional sulit untuk dilatih karena gradien yang menghilang (*vanishing gradient*) [13]. Masalah gradien berasal dari perkalian terus menerus dalam *backpropagation through time* (BPTT) [14] dengan meningkatnya kebutuhan untuk mempelajari ketergantungan jangka panjang dan skala yang besar. Dari situ, muncul variasi