

ABSTRAK

Fine-Tuning LLM dengan QLoRA untuk Chatbot Berbasis Pengetahuan Organisasi di PT. United Tractors Tbk

Oleh

Annisa Qurrota A'yun

Akses cepat terhadap informasi perusahaan penting untuk meningkatkan efektivitas kerja, mendukung pengambilan keputusan, dan mendorong inovasi. Di United Tractors, karyawan masih menghadapi kesulitan menemukan informasi relevan karena pengetahuan perusahaan tersebar di berbagai sumber. Untuk menjawab tantangan tersebut, penelitian ini menerapkan *Large Language Model* (LLM) yang di-*fine-tune* pada domain khusus dan mengevaluasi performanya dibandingkan dengan *base model*. Proses *fine-tuning* dilakukan menggunakan dataset *annual report* tahun 2023 dalam bahasa Indonesia. Seluruh data digunakan untuk pelatihan, sedangkan 20% di antaranya dipilih secara acak dan diparafrase untuk evaluasi. Model yang digunakan adalah Llama 3.1 8B dengan pendekatan QLoRA, dijalankan di Google Colab berbasis GPU T4. Proses dilakukan secara iteratif sebanyak delapan belas kali untuk memperoleh konfigurasi terbaik. Evaluasi dilakukan melalui dua pendekatan, yaitu penilaian subjektif berbasis GPT-4 dengan metrik keramahan, keringkasan, kebergunaan, dan akurasi serta pengukuran objektif menggunakan BERTScore (precision, recall, dan F1-score). Hasil menunjukkan bahwa model hasil fine-tuning mengalami peningkatan performa pada hampir seluruh metrik, dengan skor rata-rata meningkat dari 2,09 menjadi 2,95. Metrik keramahan dan keringkasan memperoleh skor tertinggi masing-

masing sebesar 4,09 dan 4, sementara kebergunaan dan akurasi memperoleh skor 3 dan 2,38. Nilai F1 BERTScore juga meningkat dari 0,659 menjadi 0,818, yang menandakan peningkatan kesamaan semantik. Temuan ini membuktikan bahwa fine-tuning efektif dalam meningkatkan kualitas respons model, khususnya dalam aspek keringkasan, keramahan, dan kesesuaian semantik dengan jawaban referensi. Namun, model masih memiliki keterbatasan dalam menghasilkan informasi faktual atau angka kuantitatif yang akurat sebagaimana tercantum dalam annual report.

Kata kunci: LLM, Chatbot, QLoRA, BERTScore, Evaluasi GPT-4

ABSTRACT

Fine-Tuning LLM with QLoRA for Organization Knowledge-Based Chatbots at PT. United Tractors Tbk

By

Annisa Qurrota A'yun

Quick access to important company information is essential for improving work effectiveness, supporting decision-making, and encouraging innovation. At United Tractors, employees still face difficulties finding relevant information because company knowledge is scattered across various sources. To address this challenge, this study applies a Large Language Model (LLM) that is fine-tuned to a specific domain and evaluates its performance compared to the base model. The fine-tuning process was carried out using the 2023 annual report dataset in Indonesian. All data was used for training, while 20% of it was randomly selected and paraphrased for evaluation. The model used was Llama 3.1 8B with the QLoRA approach, run on Google Colab based on GPU T4. The process was carried out iteratively eighteen times to obtain the best configuration.

The evaluation was conducted using two approaches, namely subjective assessment based on GPT-4 with metrics of friendliness, conciseness, usefulness, and accuracy, as well as objective measurement using BERTScore (precision, recall, and F1-score). The results show that the fine-tuned model experienced improved performance across almost all metrics, with the average score increasing from 2.09 to 2.95. The friendliness and conciseness metrics received the highest scores of 4.09 and 4, respectively, while usefulness and accuracy received scores of 3 and 2.38.

The BERTScore F1 value also increased from 0.659 to 0.818, indicating an increase in semantic similarity. These findings prove that fine-tuning is effective in improving the quality of model responses, particularly in terms of conciseness, friendliness, and semantic similarity to the reference answers. However, the model still has limitations in generating accurate factual information or quantitative figures as stated in the annual report.

Keywords: LLM, Chatbot, QLoRA, BERTScore, GPT-4 Evaluation