

## BAB V

### KESIMPULAN, IMPLIKASI, DAN REKOMENDASI

#### 5.1 Kesimpulan

Pada bagian ini akan disajikan kesimpulan temuan utama dari penelitian ini mencakup poin-poin penting yang diperoleh dari analisis yang dilakukan dan bagaimana hasil-hasil tersebut menjawab rumusan masalah penelitian yang telah disebutkan.

1. Evaluasi kinerja implementasi *fine-tuning* dan *base model* GPT-3.5 dalam menghasilkan intervensi afektif yang relevan dan efisien secara real-time sesuai emosi pelajar selama pembelajaran daring sinkronis menggunakan dua metode evaluasi yang mana Penggunaan Token dan BERTScore.
  - a. Hasil evaluasi menunjukkan bahwa proses *fine-tuning* pada *base model* GPT-3.5 memberikan penurunan yang signifikan dalam jumlah rata-rata token yang digunakan, terutama pada prompt tokens dan total tokens. Setelah *fine-tuning*, penggunaan prompt token rata-rata menurun menjadi 125.51 tokens pada *fine-tuned* model GPT-3.5. Selain itu, total tokens rata-rata yang digunakan berkurang menjadi 173.23 pada *fine-tuned* model GPT-3.5. Meskipun completion tokens meningkat menjadi 47.72 dapat mengindikasikan model memberikan respons yang detail dan relevan sehingga token lebih tinggi. Hal ini mencerminkan model model GPT-3.5 yang telah di *fine-tuning* dapat mengoptimasi dalam pembentukan *prompt* sehingga meningkatkan efisiensi penggunaan sumber daya dibandingkan *base model* GPT-3.5.
  - b. Hasil evaluasi dengan metrik BERTScore menunjukkan bahwa proses *fine-tuning* pada *base model* GPT-3.5 memberikan peningkatan signifikan pada metrik evaluasi BERTScore, yaitu Precision, Recall, dan F1-Score. Setelah *fine-tuning* nilai rata-rata atau *mean*, Precision meningkat menjadi 0.78, Recall meningkat menjadi 0.78, dan F1 Score meningkat menjadi 0.78. Hal ini membuat *fine-tuned* model GPT-3.5 tidak hanya lebih efisien dalam penggunaan token, tetapi juga

meningkatkan kualitas response dibandingkan *base* model GPT-3.5. Selain itu, evaluasi selama fine-tuning menunjukkan training loss yang fluktuatif, tetapi menurun hingga mencapai nilai 0.8341 yang mengindikasikan bahwa model berhasil mempelajari detail data pelatihan dengan baik dan memiliki kemampuan prediksi yang semakin meningkat seiring dengan berkurangnya loss.

2. Implementasi *fine-tuning* dan *base model* GPT-3.5 dalam menghasilkan intervensi afektif yang relevan dan efisien secara real-time sesuai emosi pelajar selama pembelajaran daring sinkronis memberikan beberapa perbandingan.
  - a. Penurunan signifikan dalam penggunaan prompt token dan total token mengindikasikan bahwa model GPT-3.5 dengan *fine-tuning* lebih hemat dalam penggunaan sumber daya komputasi dibandingkan *base* model GPT-3.5. Efisiensi ini penting dalam skenario pembelajaran daring dimana sumber daya terbatas bisa menjadi penghalang dalam menyediakan layanan pembelajaran yang responsif.
  - b. Dengan peningkatan pada metrik BERTScore, seperti Precision, Recall, dan F1-Score dibandingkan *base* model GPT-3.5 menunjukkan kemampuan yang lebih baik pada *fine-tuned* model GPT-3.5 dalam menghasilkan respons yang tidak hanya efisien tetapi juga relevan. Respons yang relevan dapat memberikan intervensi afektif sesuai emosi pelajar selama pembelajaran daring sinkronis.

## 5.2 Implikasi

Pada bagian ini akan membahas dampak dari hasil penelitian terhadap bidang terkait, termasuk kontribusi terhadap teori yang ada dan bagaimana hasil tersebut dapat diterapkan dalam praktik.

1. Hasil dari penelitian ini menunjukkan bahwa implementasi dengan *fine-tuning* model GPT-3.5 dapat menghasilkan intervensi afektif yang relevan dan efisien sesuai emosi pelajar secara *real-time* selama pembelajaran daring sinkronis dibandingkan *base* model GPT-3.5. Hal ini membuka peluang besar untuk personalisasi pembelajaran yang dapat menyesuaikan keadaan

emosional pelajar dengan pembelajaran sehingga potensial meningkatkan keterlibatan dan efektivitas pembelajaran.

### 5.3 Rekomendasi

Pada bagian ini akan disajikan beberapa rekomendasi yang diharapkan dapat menjadi pedoman untuk penelitian lebih lanjut atau untuk implementasi dalam praktik berdasarkan hasil temuan penelitian ini.

1. Menggunakan berbagai model lain terutama dalam OpenAI juga, seperti GPT-4, GPT-4o, GPT-4o-mini, o1, dan o1-mini. Implementasi ini bisa memberikan perspektif baru mengenai kapabilitas dan keefektifitasan model-model lain.
2. Meneliti berbagai teknik prompting lain yang bisa lebih spesifik dan disesuaikan. Penelitian ini diharapkan dapat meningkatkan relevansi dan akurasi output yang dihasilkan oleh model-model tersebut.
3. Melakukan eksperimen distribusi data dengan proporsi yang berbeda. Eksperimen dalam variasi distribusi data dapat memberikan wawasan lebih lanjut tentang bagaimana model merespons perubahan dalam input data dan bagaimana ini mempengaruhi performa model secara keseluruhan.
4. Melakukan optimasi dengan berbagai *hyperparameter* dalam proses *fine-tuning*, seperti *learning rate*, *batch size*, dan *epoch* terhadap model. Hal ini untuk mengidentifikasi konfigurasi optimal yang menghasilkan hasil terbaik untuk implementasi spesifik.
5. Melakukan eksplorasi dengan berbagai metode *fine-tuning*, seperti Parameter-Efficient Fine-Tuning (PEFT), Low-Rank Adaptation (LoRA), Prompt Tuning, dan Adapter Tuning. Eksplorasi ini bertujuan untuk membandingkan efektivitas masing-masing metode dalam meningkatkan performa model pada berbagai kasus penggunaan dan memastikan model dapat beradaptasi dengan kebutuhan yang lebih spesifik.