

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Hasil benchmarking menunjukkan bahwa Ajaib memiliki proporsi sentimen positif lebih tinggi dibandingkan Bibit pada ketiga aspek yang dianalisis. Ajaib memperoleh sentimen positif sebesar 47,9% pada aspek Feature & Functionality, 57,9% pada App Interface, dan 52,4% pada User Satisfaction. Sementara itu, Bibit memperoleh sentimen positif sebesar 29,0% pada Feature & Functionality, 53,0% pada App Interface, dan 44,8% pada User Satisfaction. Perbandingan ini dilakukan berdasarkan proporsi sentimen agar hasil lebih adil karena jumlah ulasan Ajaib dan Bibit tidak seimbang. Selain itu, Bibit memperoleh proporsi sentimen positif terendah pada aspek Feature & Functionality sebesar 29,0% dan proporsi sentimen negatif tertinggi pada aspek yang sama sebesar 61,1%. Temuan ini menunjukkan bahwa aspek Feature & Functionality pada Bibit menjadi bagian yang paling perlu diperhatikan dalam hasil benchmarking.
2. Hasil pemodelan menunjukkan bahwa IndoBERT-Base memberikan performa lebih baik dibandingkan IndoBERT-Lite pada seluruh aspek. Pada tahap hyperparameter tuning, IndoBERT-Lite memperoleh nilai Val F1 tertinggi sebesar 0,4278 pada aspek Feature & Functionality, 0,4121 pada App Interface, dan 0,4564 pada User Satisfaction. Sementara itu, IndoBERT-Base memperoleh nilai Val F1 yang lebih tinggi, yaitu 0,7982 pada Feature & Functionality, 0,7004 pada App Interface, dan 0,8370 pada User Satisfaction. Berdasarkan hasil tersebut, IndoBERT-Base dipilih sebagai model utama untuk evaluasi akhir. Pada test set, IndoBERT-Base memperoleh F1-score sebesar 0,7904 pada Feature & Functionality, 0,6778 pada App Interface, dan 0,8287 pada User Satisfaction. Aspek User Satisfaction menjadi aspek dengan performa terbaik, sedangkan App Interface memperoleh performa terendah yang dapat dipengaruhi oleh jumlah data efektif yang lebih sedikit, distribusi kelas yang lebih tidak

seimbang, serta adanya tumpang tindih makna dengan aspek User Satisfaction.

3. Hasil analisis sentimen berbasis aspek berhasil diimplementasikan dalam bentuk website ReksaBench. Website ini dikembangkan menggunakan framework Flask, diintegrasikan dengan model IndoBERT-Base hasil fine-tuning, dan di-hosting melalui Hugging Face Spaces agar dapat diakses. ReksaBench mampu menyajikan informasi performa model, visualisasi distribusi sentimen, perbandingan benchmarking antara aplikasi Bibit dan Ajaib, kata kunci dominan, contoh ulasan asli, serta fitur tambahan yakni prediksi sentimen secara langsung melalui input teks tunggal maupun unggahan file CSV.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, saran yang dapat diberikan adalah sebagai berikut:

1. Objek penelitian dapat diperluas dengan menambahkan aplikasi investasi lain, seperti Bareksa, Makmur, atau aplikasi reksa dana lainnya. Penambahan objek penelitian dapat menghasilkan perbandingan yang lebih luas dan representatif.
2. Mengingat aspek App Interface memperoleh performa paling rendah, penelitian berikutnya dapat mencoba strategi tambahan untuk menangani data yang ambigu dan tidak seimbang, seperti focal loss, oversampling, data augmentation, penyempurnaan pedoman anotasi, serta penambahan data pada kelas minoritas. Strategi ini diharapkan dapat membantu model mengenali pola sentimen pada aspek App Interface secara lebih baik, terutama karena aspek tersebut lebih jarang dibahas secara eksplisit dan memiliki tumpang tindih makna dengan aspek User Satisfaction.
3. Penelitian selanjutnya juga dapat mengembangkan proses penentuan aspek menggunakan metode topic modeling seperti LDA atau BERTopic. Metode tersebut dapat digunakan untuk mengidentifikasi topik yang muncul secara alami dari ulasan pengguna dan menemukan kandidat aspek baru.
4. Pengembangan fitur prediksi pada website dapat dilakukan dengan menambahkan mekanisme deteksi aspek sebelum proses klasifikasi

sentimen. Pada penelitian ini, sistem masih menampilkan hasil prediksi sentimen untuk tiga aspek yang telah ditentukan, sehingga penelitian selanjutnya dapat mengembangkan model tambahan untuk mendeteksi apakah suatu ulasan benar-benar membahas aspek *Feature & Functionality*, *App Interface*, atau *User Satisfaction*. Dengan adanya deteksi aspek, sistem dapat mengurangi kemungkinan pemberian sentimen pada aspek yang sebenarnya tidak dibahas oleh pengguna.

5. Penelitian selanjutnya dapat mempertimbangkan kembali struktur hubungan antar aspek yang digunakan, khususnya antara *Feature & Functionality*, *App Interface*, dan *User Satisfaction*. Hal ini karena *User Satisfaction* merupakan aspek yang bersifat lebih luas dan merepresentasikan tingkat kepuasan pengguna secara keseluruhan, yang dipengaruhi oleh pengalaman mereka dalam menggunakan fitur, fungsi, maupun antarmuka aplikasi. Oleh karena itu, penempatan ketiga aspek tersebut pada level yang sama berpotensi menimbulkan tumpang tindih secara konseptual. Sebagai alternatif, aspek *Feature & Functionality* dan *App Interface* dapat diposisikan sebagai subaspek atau faktor yang memengaruhi *User Satisfaction*, sehingga hubungan antar aspek menjadi lebih jelas dan tidak menimbulkan bias dalam proses analisis.