

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka dapat ditarik beberapa kesimpulan yang menjawab rumusan masalah sebagai berikut:

1. Hasil ekstraksi aspek menggunakan *Latent Dirichlet Allocation* (LDA) mendapatkan nilai *coherence score* tertinggi sebesar 0,4612 pada skenario *stemmed* dengan tiga topik. Ketiga topik tersebut diinterpretasikan menjadi aspek yang terdiri atas *Order and Fulfillment Process*, *Product Quality and Outlet Experience*, dan *Customer Digital Experience*, yang masing-masing menggambarkan proses pemesanan dan pemenuhan pesanan, kualitas produk dan pengalaman di *outlet*, serta pengalaman penggunaan fitur digital pada aplikasi.
2. Kinerja IndoBERT dan IndoRoBERTa dalam melakukan klasifikasi emosi *multi-label* dan sentimen berbasis aspek menunjukkan hasil yang berdekatan dengan pola yang bervariasi pada setiap tahap evaluasi. Pada k-fold cross validation, IndoRoBERTa memperoleh rata-rata *macro F1-score* sebesar 0,7105, sedikit lebih tinggi dibandingkan IndoBERT yang memperoleh 0,7097, dengan selisih sebesar 0,0008. Namun, pada komparasi menggunakan data uji, hasil tersebut berbalik, di mana IndoBERT memperoleh rata-rata *macro F1-score* sebesar 0,7334, lebih tinggi dibandingkan IndoRoBERTa yang memperoleh 0,7289. Pola serupa juga terlihat pada tingkat aspek, IndoBERT unggul pada aspek *Order and Fulfillment Process* dan *Product Quality and Outlet Experience*, sedangkan IndoRoBERTa unggul pada aspek *Customer Digital Experience*. Pada metrik pendukung, IndoBERT menghasilkan rata-rata *macro recall* dan *accuracy* yang lebih tinggi, sementara IndoRoBERTa menghasilkan rata-rata *macro precision* yang lebih tinggi serta *total error* yang lebih rendah. Temuan ini menunjukkan bahwa kedua model memiliki keunggulan pada metrik dan tahap evaluasi yang berbeda, sehingga tidak ada satu model yang secara konsisten unggul pada seluruh kondisi pengujian.

3. Berdasarkan hasil komparasi kinerja model, IndoBERT dengan *class weighting* dan pembagian data 70:30 menunjukkan performa terbaik dalam mengklasifikasikan emosi dan sentimen berbasis aspek pada ulasan pengguna aplikasi kedai kopi. Model tersebut memperoleh rata-rata *macro F1-score* sebesar 0,7334, lebih tinggi sebesar 0,0045 dibandingkan IndoRoBERTa yang memperoleh 0,7289. Hasil ini sesuai dengan penggunaan *macro F1-score* sebagai metrik utama dalam penentuan model terbaik.
4. IndoBERT sebagai model terbaik diimplementasikan menggunakan Flask menjadi sistem analisis emosi *multi-label* dan sentimen berbasis aspek dalam bentuk *website*. Sistem dapat menerima *input* berupa satu ulasan melalui fitur Prediksi Teks maupun beberapa ulasan melalui fitur Prediksi File dalam format CSV. Setiap ulasan yang dimasukkan terlebih dahulu diidentifikasi ke dalam aspek yang telah ditetapkan berdasarkan hasil interpretasi LDA, kemudian model IndoBERT memprediksi emosi *multi-label* dan sentimen sesuai dengan aspek yang teridentifikasi. Hasil analisis ditampilkan dalam bentuk tabel dan visualisasi. Dengan demikian, sistem dapat digunakan untuk menganalisis emosi *multi-label* dan sentimen pada ulasan aplikasi kedai kopi secara otomatis berdasarkan aspek yang dibahas.

5.2 Saran

Berdasarkan hasil penelitian, beberapa saran yang dapat diberikan untuk pengembangan selanjutnya adalah:

1. Beberapa label emosi dengan jumlah kemunculan terbatas, seperti *disgust* dan *fear*, masih memperoleh nilai *F1-score* yang rendah. Penelitian selanjutnya dapat menambahkan jumlah dan variasi data pada label tersebut melalui pengumpulan ulasan baru atau augmentasi data. Penambahan data dapat memberikan lebih banyak contoh bagi model dalam mempelajari pola setiap emosi.
2. Penelitian ini menggunakan *class weighting* dan *focal loss* untuk menangani ketidakseimbangan label. Penelitian selanjutnya dapat menguji metode lain, seperti *Asymmetric Loss*, untuk mengetahui kinerjanya dalam menangani label emosi dengan jumlah kemunculan terbatas.

3. *Hyperparameter tuning* pada penelitian selanjutnya dapat dilakukan secara terpisah pada setiap aspek dan skenario pembagian data apabila sumber daya komputasi memadai. Pada penelitian ini, *tuning* dilakukan pada aspek CDE dengan skenario 80:20, kemudian konfigurasi terbaik diterapkan pada seluruh skenario. *Tuning* secara terpisah berpotensi menghasilkan konfigurasi yang lebih sesuai dengan karakteristik data masing-masing aspek dan skenario.
4. Penelitian selanjutnya dapat membandingkan IndoBERT dan IndoRoBERTa dengan model *transformer* multibahasa, seperti XLM-RoBERTa atau mBERT. Pengujian tersebut dapat dilakukan untuk mengetahui kinerja model pada ulasan berbahasa Indonesia yang memuat istilah berbahasa Inggris.
5. Ekstraksi aspek menggunakan LDA pada penelitian ini masih bekerja secara probabilistik berbasis *bag-of-words* dan memerlukan skenario *stemming* untuk mencapai *coherence score* tertinggi. Penelitian selanjutnya dapat menggunakan BERTopic yang memanfaatkan representasi kontekstual dari model *transformer* sebagai metode ekstraksi aspek. Pendekatan ini berpotensi menghasilkan topik yang lebih representatif sehingga kualitas pelabelan aspek dapat lebih akurat.
6. Penelitian ini belum mendeteksi ulasan spam atau ulasan yang berasal dari akun bot. Penelitian selanjutnya dapat menambahkan fitur deteksi spam atau bot pada sistem yang dibangun, untuk meningkatkan kualitas ulasan yang diproses.