

BAB V

PENUTUP

5.1 Kesimpulan

Penelitian ini melakukan analisis sentimen berbasis aspek pada ulasan hotel menggunakan pendekatan *Multilabel Aspect-Based Sentiment Analysis (ABSA)* dengan membandingkan berbagai kombinasi arsitektur *deep learning*, *word embedding*, dan teknik *preprocessing*. Berdasarkan hasil penelitian yang telah dilakukan, diperoleh beberapa kesimpulan sebagai berikut:

1. Distribusi label dari 7.821 ulasan menunjukkan perbedaan sentimen pada setiap aspek perhotelan. Aspek *Room* menjadi aspek yang paling banyak menerima keluhan, dengan 1.812 ulasan negatif dan 1.595 ulasan positif. Hal ini mengindikasikan bahwa kondisi kamar, seperti kebersihan, ukuran, dan fasilitas, masih menjadi perhatian utama pengunjung. Sebaliknya, aspek *Location* memperoleh penilaian paling positif, dengan 2.039 ulasan positif dan 603 ulasan negatif, yang menunjukkan tingginya kepuasan terhadap lokasi hotel. Aspek *Service* juga didominasi sentimen positif, yaitu 1.784 ulasan positif berbanding 995 ulasan negatif, sedangkan aspek *Hotel* mencatat 2.001 ulasan positif dan 1.280 ulasan negatif. Secara umum, lokasi hotel menjadi keunggulan utama, sementara aspek kamar masih menjadi sumber keluhan terbesar pengunjung.
2. Komparasi delapan skenario pada setiap model klasifikasi menunjukkan bahwa tidak ada satu kombinasi model yang selalu unggul pada seluruh task, sehingga konfigurasi terbaik sangat bergantung pada karakteristik masing-masing task klasifikasi. Dari sisi arsitektur, BiLSTM memberikan hasil terbaik pada model klasifikasi aspek multilabel karena mampu memahami konteks kalimat dua arah dan relasi antar aspek dalam satu ulasan. Sebaliknya, CNN lebih unggul pada tiga dari empat model klasifikasi sentimen, yaitu aspek *Room*, *Location*, dan *Service*, karena lebih efektif menangkap pola kata kunci lokal yang umum muncul pada opini sentimen. Pada model klasifikasi sentimen aspek *Hotel*, tidak ditemukan dominasi

arsitektur yang konsisten, meskipun model terbaik menggunakan arsitektur BiLSTM. Dari sisi *word embedding*, Word2Vec secara konsisten mengungguli FastText pada model klasifikasi sentimen aspek *Hotel*, *Location*, dan *Service*, sedangkan pada klasifikasi aspek multilabel dan sentimen *Room* keduanya menunjukkan performa yang relatif setara. Hal ini menunjukkan bahwa Word2Vec lebih mampu merepresentasikan kosakata kontekstual pada beberapa aspek, sementara keunggulan FastText dalam menangani variasi kata tidak terlalu berpengaruh setelah proses normalisasi teks dilakukan. Dari sisi *preprocessing*, pengaruh *stemming* tidak konsisten. *Stemming* meningkatkan performa pada beberapa model, terutama klasifikasi aspek multilabel dan klasifikasi sentimen *Service*, tetapi menurunkan performa pada model klasifikasi sentimen aspek *Hotel* dan *Location* karena sebagian informasi kata yang relevan ikut hilang.

3. Penelitian ini menghasilkan lima model terbaik melalui pengujian delapan kombinasi skenario menggunakan *5-Fold Cross Validation*. Model tersebut terdiri atas satu model klasifikasi aspek multilabel dan empat model klasifikasi sentimen per aspek, yang dipilih berdasarkan nilai *F1 Macro* tertinggi dan performa terbaik. Berdasarkan hasil evaluasi performa, diperoleh kombinasi model terbaik pada masing-masing task sebagai berikut:
 - a. Model Klasifikasi Aspek Multilabel terbaik adalah model dari Skenario 8 (BiLSTM + FastText + Stemming) dengan *F1 Macro* 97.02%, *Hamming Loss* 2.38%, dan *Subset Accuracy* 91.55%.
 - b. Model Klasifikasi Sentimen Aspek *Room* terbaik adalah model dari Skenario 4 (CNN + FastText + Stemming) dengan *F1 Macro* 97.84% dan AUC-PR 0.9965.
 - c. Model Klasifikasi Sentimen Aspek *Hotel* terbaik adalah model dari Skenario 5 (BiLSTM + Word2Vec + NoStemming) dengan *F1 Macro* 96.56% dan AUC-PR 0.9891.
 - d. Model Klasifikasi Sentimen Aspek *Location* terbaik adalah model dari Skenario 1 (CNN + Word2Vec + NoStemming) dengan *F1 Macro* 97.37% dan AUC-PR 0.9864.

- e. Model Klasifikasi Sentimen Aspek *Service* terbaik adalah model dari Skenario 2 (CNN + Word2Vec + Stemming) dengan *F1 Macro* 97.96% dan AUC-PR 0.9952.

Pengujian pada data uji (test set) yang tidak terlibat selama pelatihan juga menunjukkan performa yang tinggi dan stabil. Model klasifikasi aspek multilabel mencapai *F1 Macro* 97.25%, *Hamming Loss* 2.17% dan *Subset Accuracy* 92.70%. Model klasifikasi sentimen aspek *Room* mencapai *F1 Macro* 98.82%, aspek *Hotel* 95.16%, aspek *Location* 96.76%, dan aspek *Service* 97.25%. Selisih performa yang relatif kecil antara validasi K-Fold dan data test menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak menunjukkan indikasi *overfitting*.

4. Lima model terbaik yang telah dipilih selanjutnya di-deploy ke dalam sistem web. Sistem web berhasil diimplementasikan menggunakan *framework* Flask dengan integrasi kelima model dalam satu pipeline otomatis. Sistem mampu melakukan analisis secara otomatis dan *end-to-end* mulai dari deteksi bahasa, penerjemahan ulasan berbahasa Inggris, *preprocessing*, hingga prediksi aspek dan sentimen. Sistem juga mendukung analisis ulasan tunggal maupun analisis batch melalui file CSV/Excel. Selain itu, sistem juga menyediakan dashboard visualisasi interaktif berupa distribusi aspek, perbandingan sentimen, tren ulasan, dan analisis *top-words* per aspek.

5.2 Saran

Berdasarkan hasil dan keterbatasan penelitian, berikut lima saran untuk pengembangan penelitian selanjutnya:

1. Penelitian ini hanya membandingkan dua arsitektur *deep learning*, yaitu CNN dan BiLSTM. Penelitian selanjutnya disarankan mengeksplorasi arsitektur yang lebih mutakhir seperti Transformer, BERT, atau IndoBERT yang memiliki kemampuan pemahaman konteks lebih baik, sehingga berpotensi mengurangi kesalahan prediksi pada kasus sarkasme, negasi ganda, dan kritik implisit.
2. Data penelitian ini hanya berasal dari satu platform, yaitu TripAdvisor, dan terbatas pada tiga wilayah. Penelitian berikutnya disarankan memperluas sumber data dari platform lain serta menambah cakupan wilayah agar model memiliki kemampuan generalisasi yang lebih baik.
3. Proses anotasi data pada penelitian ini masih dilakukan secara manual oleh tiga anotator. Penelitian selanjutnya dapat memanfaatkan teknik *active learning* atau *large language model* sebagai anotator otomatis dengan agar proses pelabelan menjadi lebih efisien.
4. Penelitian ini menggunakan pendekatan translasi ulasan berbahasa Inggris ke Bahasa Indonesia sebelum proses klasifikasi. Penelitian selanjutnya dapat mengeksplorasi penggunaan model multilingual secara langsung, seperti mBERT, XLM-R, atau IndoXLM, sehingga teks multibahasa dapat diproses tanpa tahap penerjemahan.
5. Sistem web yang dikembangkan masih bersifat statis dan memerlukan unggahan file secara manual. Penelitian selanjutnya dapat mengembangkan sistem menjadi dashboard terintegrasi berbasis real-time melalui API platform ulasan, sehingga pengelola hotel dapat memantau sentimen pelanggan secara otomatis dan berkelanjutan.