

BAB V

PENUTUP

5.1. Kesimpulan

Penelitian ini bertujuan untuk menganalisis sentimen wisatawan terhadap aspek-aspek tempat wisata populer di Jawa Timur melalui ulasan di platform digital seperti Tripadvisor dan Google Maps menggunakan pendekatan Aspect-Based Sentiment Analysis (ABSA) serta mengetahui skenario terbaik untuk implementasi CNN pada ABSA tersebut. Berdasarkan hasil analisis, dapat disimpulkan beberapa hal sebagai berikut:

1. Distribusi sentimen menunjukkan bahwa aspek attraction paling banyak mendapat ulasan positif (6401), menandakan daya tarik wisata dinilai baik oleh pengunjung. Aspek amenities juga lebih banyak mendapat ulasan positif (2015) dibanding negatif (802), mengindikasikan fasilitas umumnya cukup memadai. Sebaliknya, pada aspek access dan price, jumlah sentimen negatif hampir setara atau sedikit lebih rendah dari positif, menunjukkan akses dan harga cenderung mendapat penilaian beragam atau kurang memuaskan.
2. Dari total 5 model yang diuji pada tiap skenario (total 12 skenario) dalam penelitian ini yang terdiri dari 1 model klasifikasi aspek secara multilabel dan 4 model klasifikasi sentimen untuk setiap aspek, dipilih masing-masing model terbaik berdasarkan nilai macro average F1-score dan ROC AUC.
 - a. Model Multilabel Aspek, F1-score = 0.8120, AUC = 0.9111 (skenario 3, Word2Vec + CNN+ stem).
 - b. Model Sentimen Attraction, F1-score = 0.7112, AUC = 0.9437 (skenario 9, IndoBERT + CNN).
 - c. Model Sentimen Amenities, F1-score = 0.8503, AUC = 0.9312 (skenario 9, IndoBERT + CNN).
 - d. Model Sentimen Access, F1-score = 0.7567, AUC = 0.8312 (skenario 9, IndoBERT + CNN).
 - e. Model Sentimen Price, F1-score = 0.8096, AUC = 0.8905 (skenario 9, IndoBERT + CNN).
3. Hasil pengujian masing-masing model terpilih yang sudah melalui hyperparameter tuning menunjukkan bahwa model memiliki performa yang stabil dan cukup baik, bahkan cenderung meningkat dari hasil validasi:

- a. Model Multilabel Aspek, F1-score = 0.83, AUC = 0.9087.
 - b. Model Sentimen Attraction, F1-score = 0.80, AUC = 0.8947.
 - c. Model Sentimen Amenities, F1-score = 0.86, AUC = 0.9328.
 - d. Model Sentimen Access, F1-score = 0.75, AUC = 0.8352.
 - e. Model Sentimen Price, F1-score = 0.83, AUC = 0.9192.
4. Penggunaan Stemming cenderung memberikan dampak positif pada model dengan Word2Vec, karena stemming menyederhanakan kata yang mirip. Sedangkan pada model dengan BERT-based, tidak konsisten.
 5. Random Oversampling (ROS) memberikan pengaruh yang tidak konsisten dan cenderung kurang efektif terutama pada model berbasis BERT dan IndoBERT.
 6. Model terbaik yang telah dipilih kemudian di-*deploy* dalam sistem berbasis web menggunakan Flask yang memungkinkan input ulasan berupa single text maupun batch dalam file CSV yang kemudian dianalisis secara otomatis untuk mendeteksi aspek dan sentimennya dengan hasil ditampilkan dalam bentuk tabel dan visualisasi pie chart sehingga memudahkan pemahaman data ulasan.

5.2. Saran

Berdasarkan hasil penelitian ini, terdapat beberapa hal yang dapat disarankan untuk pengembangan lebih lanjut dalam penelitian serupa. Saran-saran berikut diberikan untuk mengatasi keterbatasan yang belum sepenuhnya terselesaikan, serta untuk meningkatkan kualitas model dan akurasi prediksi di masa depan:

1. Disarankan untuk mengeksplorasi metode penanganan ketidakseimbangan kelas yang lain selain Random Oversampling (ROS), seperti SMOTE, atau penggunaan class weight, atau mungkin dengan data augmentasi.
2. Menambahkan dan menyeimbangkan data ulasan yang merata tiap lokasi agar tidak menimbulkan potensi bias geografis dan model bisa menangkap pola yang lebih luas tiap lokasinya.
3. Penambahan data yang lebih banyak agar model deep learning seperti CNN dapat belajar lebih banyak pola ulasan.