

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian, *Algoritma DistilBERT* terbukti dapat diterapkan secara efektif dan efisien dalam melakukan analisis sentimen terhadap ulasan hotel di Indonesia dengan tiga kategori sentimen, yaitu positif, negatif, dan netral. Model ini juga menunjukkan performa yang kompetitif jika dibandingkan dengan model-model terkini lainnya. Dari kelima hotel yang diambil Hasil evaluasi paling tinggi berada di hotel *galery prawirotaman* menunjukkan bahwa *DistilBERT* mampu menghasilkan *akurasi* sebesar 96%, *precision* 96%, *recall* 96%, dan *f1-score* 96%, yang mengindikasikan kemampuan model dalam menangani analisis sentimen berbahasa Indonesia secara andal.

Penerapan berbagai tahapan *pre-processing* data seperti *case folding*, penghapusan spasi ganda (*remove multi-space*), *filtering*, pembersihan tanda pagar (*clean hashtag*), penghapusan tanda baca (*remove punctuation*), pembersihan emoji, serta *stemming* dan *stopword*, berkontribusi terhadap peningkatan kualitas data dan efisiensi proses pelatihan model. Selain itu, penerapan metode *Random Oversampling*(ROS) dalam proses penyeimbangan kelas pada dataset turut meningkatkan kemampuan model dalam mengklasifikasikan data dengan distribusi kelas yang beragam. Kombinasi teknik *pre-processing*, penyeimbangan data, memainkan peran penting dalam keberhasilan implementasi *DistilBERT* pada analisis sentimen ini.

5.2 Saran

1. Menggunakan dataset dengan jumlah kategori yang lebih beragam atau melakukan penyeimbangan distribusi kategori pada dataset. Bertujuan agar *model* tidak hanya belajar dari pola spesifik pada data pelatihan, tetapi juga mampu mengenali dan memahami pola umum yang terdapat pada berbagai jenis ulasan lainnya.
2. mempertimbangkan penggunaan teknik penyeimbangan data lain, seperti SMOTE (*Synthetic Minority Over-sampling Technique*), yang tidak hanya menduplikasi data seperti pada metode ROS, tetapi juga menghasilkan

sampel sintetis baru berdasarkan kemiripan antar data, sehingga dapat meningkatkan keragaman dan representasi kelas minoritas dalam dataset.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu di bidang machine learning, khususnya dalam penerapan analisis sentimen. Selain itu, hasil dari penelitian ini juga diharapkan dapat membantu pihak industri perhotelan dalam memahami opini dan persepsi pelanggan secara lebih efektif.

