

BAB 5

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan, pelatihan model untuk memprediksi tipe kepribadian MBTI seseorang berdasarkan konten media sosial dengan menggunakan algoritma Support Vector Machines telah selesai diimplementasikan. Proses pengimplementasian dilakukan pada *server* Google Collaboratory dengan bahasa pemrograman Python. *Dataset* yang digunakan diperoleh dari Kaggle (<https://www.kaggle.com/datasets/datasnaek/mbti-type>) sebanyak 8675 baris data. Pelatihan model dilakukan dengan menerapkan *text preprocessing*, *feature extraction* dengan TF-IDF, *oversampling* dengan SMOTE, serta menerapkan algoritma Support Vector Machine pada model. Pengujian yang dilakukan membuktikan bahwa dengan melakukan *oversampling* dengan SMOTE dapat meningkatkan akurasi pada model. Nilai akurasi yang diperoleh pada pelatihan model yaitu sebesar 93,42%.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, beberapa saran yang dapat dipertimbangkan untuk perkembangan penelitian serupa di masa mendatang antara lain sebagai berikut.

1. *Data splitting* dilakukan dengan Stratified Shuffle Split agar pada data latih dan data uji yang dipisah akan memuat jumlah data yang sama pada setiap kelas sehingga model dapat mengenali data dengan proporsi yang seimbang yang kemungkinan dapat meningkatkan akurasi model.
2. Melakukan *preprocessing* pada tahap *lemmatization* dengan *pos tagging* yang dapat menghasilkan akar kata yang lebih akurat sehingga mungkin dapat meningkatkan akurasi pelatihan model.