

BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil dari implementasi dan hasil uji coba sistem, dapat disimpulkan bahwa implementasi metode Unsupervised Data Augmentation untuk klasifikasi teks *hate speech* telah selesai dibuat. Hasil uji coba memberikan performa teroptimal dengan model UDA yang dilatih menggunakan *supervised training data* sebanyak 10% dari keseluruhan data dan *dataset* yang melalui proses *stemming* dan penghilangan *stopword*. Dalam penelitian ini, model teroptimal tersebut memiliki nilai *accuracy* sebesar 78,8%, *precision* sebesar 79,1%, *recall* sebesar 77,9%, dan *F1 score* sebesar 77%.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk penelitian selanjutnya, yaitu sebagai berikut.

1. Mencoba untuk melakukan penyeleksian terhadap *supervised training data* yang digunakan untuk melatih model UDA secara manual ataupun dengan menggunakan metode tertentu.
2. Mencoba penerapan algoritma yang mirip dengan Training Signal Annealing (TSA) dalam proses *confidence-based masking*.
3. Mencoba penggunaan BERT-Large model ataupun penerapan *fine-tuning* BERT pada *unsupervised training data*, sehingga proses *pre-training* dari

BERT model dilanjutkan kembali agar dapat disesuaikan dengan fitur-fitur *unsupervised training data* dalam *dataset*.