

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil penelitian yang telah dilakukan mengenai klasifikasi konten seksual pada lirik lagu menggunakan algoritma *Extra Trees Classifier (ETC)*, *Long Short-Term Memory (LSTM)*, serta pendekatan *soft voting ensemble*, diperoleh beberapa simpulan sebagai berikut:

1. Klasifikasi lirik lagu berdasarkan konten seksual menggunakan metode *soft voting* dengan menggabungkan metode *Extra Trees Classifier* dan *Long Short-Term Memory* berhasil dilakukan. Proses klasifikasi dimulai dengan tahapan *data preprocessing* seperti *cleaning*, *tokenizing*, *normalization*, *stopword removal*, dan *stemming* untuk menghasilkan data bersih yang siap diklasifikasikan. Setelah itu, *dataset* akan melalui proses *feature extraction* dengan *TF-IDF*, *Doc2Vec*, dan *Word2Vec*. *Soft voting* yang dilakukan dipisahkan menjadi dua, yaitu *ETC + TF-IDF* dengan *LSTM + Word2Vec* dan *ETC + Doc2Vec* dengan *LSTM + Word2Vec*.
2. Hasil evaluasi performa dari metode *soft voting* memiliki tingkat akurasi yang lebih baik dibandingkan model individual *Extra Trees Classifier* dan *Long Short-Term Memory*. Model *soft voting ETC TF-IDF + LSTM Word2Vec* menghasilkan akurasi sebesar 90,1%, dengan *precision* 49,4%, *recall* 88,3%, dan *f1-score* 63,1%. Sementara itu, kombinasi *ETC Doc2Vec + LSTM Word2Vec* menunjukkan performa terbaik dengan akurasi 98,5%, *precision* 88,0%, *recall* 96,6%, dan *f1-score* 92,2%. Dibandingkan dengan model *LSTM* yang memiliki akurasi sebesar 90,1%, dengan *precision* 49,0%, *recall* 71,8%, dan *f1-score* 58,2% dan model *ETC* yang memiliki akurasi masing-masing sebesar 94,1% dan 90,8%, dengan *precision* masing-masing sebesar 78,6% dan 100%. Namun, nilai *recall* yang dihasilkan untuk kelas eksplisit masih rendah, yaitu 53,4% dan 3,9%. Hasil ini membuktikan bahwa penggunaan pendekatan *ensemble learning* mampu menggabungkan kekuatan masing-masing algoritma untuk meningkatkan generalisasi model dalam mengidentifikasi konten seksual dalam lirik lagu secara lebih efektif.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk penelitian selanjutnya sebagai berikut:

1. Dapat dipertimbangkan penggunaan algoritma klasifikasi lain seperti *Random Forest*, *XGBoost*, atau model berbasis *transformer* (seperti *BERT*) untuk mengevaluasi performa lebih lanjut terhadap tugas klasifikasi teks ini.
2. Perlu dilakukan analisis lebih mendalam terhadap kata-kata kunci atau fitur yang paling berpengaruh dalam mendeteksi konten seksual pada lirik lagu, guna mendukung *interpretability* hasil klasifikasi.

