BAB 5
SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan penelitian yang telah dilakukan, dengan ini disimpulkan:

1. Model deteksi komentar pelecehan seksual verbal berhasil diimplementasikan melalui serangkaian tahapan sistematis. Proses ini diawali dengan *data collection*, *data labelling* manual, *random undersampling* untuk mengatasi kelas tidak seimbang, *text preprocessing*, *feature extraction* menggunakan TF-IDF, dan pembagian data (80:20). 80% *data training* digunakan untuk pelatihan model *Random Forest* yang dimulai dengan menggunakan parameter *default* sebagai acuan *baseline*, yang kemudian dilanjutkan dengan proses *hyperparameter tuning* menggunakan *GridSearchCV* untuk memperoleh performa terbaik dari algoritma tersebut. Terakhir, kedua hasil pemodelan tersebut dievaluasi dan dibandingkan performanya menggunakan 20% *data testing* untuk menentukan konfigurasi yang menghasilkan model terbaik.

2. Performa terbaik dari algoritma *Random Forest* dalam mendeteksi komentar pelecehan seksual verbal, yang ditemukan melalui proses *hyperparameter tuning*, menunjukkan hasil yang sangat efektif. Model ini berhasil mencapai akurasi keseluruhan sebesar 87,50%, yang diperoleh melalui konfigurasi parameter terbaik yaitu `n_estimators=300` dan `min_samples_leaf=2`. Secara spesifik, performa pada kelas *Non Sexual Harassment* (label 0) mencapai *precision* 93,15%, *recall* 82,93%, dan *F1-score* 87,74%. Sementara itu, untuk kelas *Sexual Harassment* (label 1), model mencapai *precision* 82,28%, *recall* 92,86%, dan *F1-score* 87,25%. Tingginya nilai *recall* pada kelas pelecehan menunjukkan kemampuan model yang andal dalam mengidentifikasi komentar pelecehan seksual.

5.2 Saran

1. Implementasi Model dalam Aplikasi atau *Website* Deteksi

   Model ini dapat diimplementasikan dalam aplikasi atau *website* deteksi komentar pelecehan seksual secara otomatis di media sosial. Dengan integrasi

ini, pengguna, terutama orang tua dan remaja, dapat menerima notifikasi jika komentar yang mereka buat atau baca mengandung pelecehan seksual. Selain itu, sistem ini dapat membantu pengelola media sosial memoderasi konten secara *real-time*, menciptakan lingkungan digital yang lebih aman dan mencegah dampak negatif bagi remaja.

2. Pengembangan Model dengan *Dataset* Lebih Besar dan Variasi Bahasa

   Untuk meningkatkan performa model, disarankan untuk mengumpulkan *dataset* yang lebih besar dan beragam, mencakup variasi bahasa dan dialek yang lebih luas. Misalnya, dengan menambahkan komentar yang mengandung bahasa gaul, slang, atau kode campur (*code-mixing*), seperti penggunaan campuran bahasa Indonesia dan bahasa Inggris (”gue lagi bosen, want to hang out?”), model akan lebih siap untuk mengenali pola bahasa yang digunakan di media sosial.

3. Eksperimen dengan Algoritma Lain untuk Deteksi Pelecehan Seksual

   Meskipun *Random Forest* telah memberikan hasil yang memuaskan, penelitian lebih lanjut dapat dilakukan dengan menguji algoritma lain yang juga terkenal dalam menangani masalah klasifikasi teks, seperti *Support Vector Machines* (SVM) atau *Naïve Bayes* (NB). Perbandingan antara algoritma ini dapat membantu untuk menemukan metode yang paling efektif dalam mendeteksi pelecehan seksual verbal di media sosial. Selain itu, teknik *ensemble learning* atau penggabungan beberapa model dapat diuji untuk meningkatkan stabilitas dan ketepatan deteksi.

4. Pengujian di Berbagai Platform dan Bahasa

   Salah satu cara untuk meningkatkan relevansi model ini adalah dengan menguji dan mengimplementasikan model pada platform sosial lainnya, seperti Facebook, Instagram, atau YouTube, yang memiliki karakteristik dan tipe data yang berbeda. Selain itu, mengembangkan model serupa dalam bahasa lain, seperti Inggris, dapat membantu untuk memperluas penerapan teknologi deteksi pelecehan seksual ini ke skala global dan mendukung keamanan digital secara lebih luas.