# BAB V

# SIMPULAN DAN SARAN

## 5.1 Simpulan

Penelitian ini berhasil membangun sebuah sistem deteksi otomatis yang mampu mengidentifikasi berbagai bentuk cyberbullying dalam teks berbahasa Indonesia dengan pendekatan multi-label classification. Sistem ini dikembangkan melalui tahapan yang sistematis, dimulai dari proses pengumpulan data melalui scraping Twitter, dilanjutkan dengan preprocessing data secara menyeluruh yang mencakup pembersihan teks, normalisasi, stemming, penghapusan rare words, serta augmentasi data untuk label minoritas. Tahapan ini diikuti oleh tokenisasi menggunakan model pralatih IndoBERT yang mampu menangkap makna semantik setiap token dalam konteks bahasa Indonesia, dan kemudian diteruskan ke dalam arsitektur Bi-LSTM untuk memproses pola urutan kata dalam dua arah. Proses ini dilengkapi dengan penggunaan Focal Loss sebagai fungsi kerugian untuk menangani ketidakseimbangan kelas, serta penerapan threshold dinamis berbasis ROC-AUC untuk meningkatkan akurasi prediksi pada label-label minor.

Model hybrid IndoBERT dan Bi-LSTM yang dibangun dalam penelitian ini menunjukkan hasil klasifikasi yang sangat baik terhadap 12 kategori cyberbullying. Model mencapai nilai F1-score mikro sebesar 0.89 dan F1-score makro sebesar 0.88 pada data validasi, dengan nilai precision dan recall yang tinggi di hampir seluruh label, baik mayoritas maupun minoritas. Bahkan untuk beberapa label dengan data terbatas seperti HS_Religion, HS_Race, dan HS_Strong, model mampu memberikan prediksi dengan nilai ROC-AUC sempurna (1.00), menunjukkan kemampuan diskriminatif model yang sangat tinggi. Evaluasi lebih lanjut melalui confusion matrix dan classification report juga menunjukkan bahwa model berhasil mengklasifikasikan berbagai jenis kekerasan verbal, baik secara eksplisit maupun implisit, tanpa mengorbankan akurasi pada label mayoritas. Penelitian ini membuktikan bahwa pendekatan hybrid yang menggabungkan kekuatan representasi semantik IndoBERT dan kemampuan urutan konteks dari Bi-LSTM ditambah dengan strategi balancing dan thresholding yang tepat mampu

menghasilkan sistem deteksi cyberbullying multi-label yang baik dan siap untuk diimplementasikan sebagai alat bantu dalam menangani perundungan daring di media sosial berbahasa Indonesia.

## 5.2 Saran

Sebagai tindak lanjut dari penelitian ini, berikut beberapa saran yang dapat dipertimbangkan :

1. **Peningkatan kualitas klasifikasi pada label dengan F1-score rendah:** Label *HS_Physical* menunjukkan F1-score yang relatif rendah (0.64), meskipun recall-nya tinggi (0.89) namun precision-nya rendah. Hal ini menunjukkan bahwa model sering salah mengklasifikasikan teks sebagai *HS_Physical*. Perlu dilakukan analisis lebih lanjut terhadap ekspresi atau kata-kata yang memicu prediksi keliru serta peningkatan variasi data pelatihan di label tersebut.
2. **Perlu ditelusuri sumber kesalahan pada prediksi HS_Physical:** Meskipun recall untuk label *HS_Physical* tinggi (0.89), precision-nya sangat rendah, menyebabkan F1-score hanya 0.64. Artinya, model sering mengira teks mengandung kekerasan fisik padahal tidak. Saran yang dapat dilakukan adalah lakukan analisis terhadap teks-teks yang salah diklasifikasi untuk melihat apakah ada kata-kata ambigu atau noise yang belum ditangani pada preprocessing.
3. **Menggunakan Kamus Stopword yang lebih lebih baik :**
   Penggunaan kamus stopword dalam penelitian ini masih memiliki keterbatasan, terutama dalam menghapus kata-kata umum seperti “itu”, “saja”, dan “dan” yang masih muncul dalam hasil WordCloud. Hal ini menunjukkan bahwa daftar stopword bawaan yang digunakan belum sepenuhnya relevan dengan karakteristik bahasa di media sosial. Oleh karena itu disarankan untuk memperluas kamus stopword dengan menggabungkan sumber dari Sastrawi, stopword informal khas media sosial, serta hasil analisis kata berfrekuensi tinggi dalam dataset. Dengan demikian, proses cleaning akan lebih optimal dan model dapat lebih fokus pada fitur-fitur yang benar-benar penting untuk deteksi kekerasan verbal.