BAB V
PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh jawaban atas rumusan masalah berdasarkan hasil implementasi dari berbagai tahap metode penelitian. Adapun kesimpulan dari penelitian ini adalah sebagai berikut.

1. Model IndoBERTweet-BiGRU berhasil diimplementasikan untuk mendeteksi teks yang mengandung unsur *cyberbullying* pada media sosial X. Proses penerapan model beberapa *hyperparameter* yang mencakup *Batch Size* sebesar 16, BiGRU *Hidden Size* sebesar 128, *Drop Out* sebesar 0.3, *Epoch* sebesar 4, *Gamma* pada *Focal Loss* sebesar 2.0, dan *Learning Rate* sebesar 2×10^{-5}. Melalui penerapan *hyperparameter* tersebut, dilakukan tiga eksperimen berbeda pada tahap *pre-processing* dan dua eksperimen pada modifikasi distribusi label. Hasil implementasi model menunjukkan adanya indikasi *overfitting* dengan selisih sekitar 0.06 antara *training accuracy* dan *validation accuracy*. Namun, pada pengujian terhadap data baru, model dapat memprediksi kelas dengan baik. Dengan demikian, model IndoBERTweet-BiGRU dapat diterapkan dengan cukup baik untuk mendeteksi *cyberbullying* pada teks media sosial berbahasa Indonesia, khususnya pada *platform* X atau Twitter, meskipun terdapat indikasi *overfitting* pada proses pelatihan model.
2. Berdasarkan hasil eksplorasi pada berbagai skenario *pre-processing* dan penanganan ketidakseimbangan kelas, diperoleh bahwa kombinasi *pre-processing* dengan skema eksperimen pertama, yaitu menerapkan token USER, menghapus tanda baca, angka, simbol, dan URL, menerapkan *lowercasing* hanya pada huruf kapital pada awal kalimat, serta tidak melakukan normalisasi dan penghapusan *stopwords* yang kemudian dikombinasikan dengan distribusi label hasil modifikasi eksperimen kedua dengan label Kata Kasar, *Sexual Harassment* dan Vulgar, *Body Shaming* dan *Hate Speech*, dan *Non-Cyberbullying* memberikan performa terbaik

dibandingkan skenario lainnya. Model pada skenario tersebut berhasil memperoleh nilai *accuracy* sebesar 0.93, *precision* sebesar 0.90, *recall* sebesar 0.90, dan *F1-score* sebesar 0.90, sehingga menunjukkan bahwa pemilihan tahapan *pre-processing* dan strategi penanganan ketidakseimbangan kelas dapat meningkatan performa klasifikasi.

3. Penerapan *Focal Loss* sebagai *loss function* mampu menangani ketidakseimbangan kelas tanpa memerlukan modifikasi distribusi label. Hal ini ditunjukkan pada performa model pada dataset asli dengan label *Kata Kasar* dan Vulgar, *Sexual Harassment, Body Shaming, Hate Speech*, dan *Non-Cyberbullying* yang menghasilkan nilai akurasi sebesar 0.93, *F1-score* sebesar 0.87, *precision* sebesar 0.85, dan *recall* sebesar 0.88 pada penerapan Focal Loss. Hasil tersebut lebih tinggi dibandingkan tanpa *Focal Loss* yang memiliki akurasi sebesar 0.88, *F1-score* sebesar 0.77, *precision* sebesar 0.80, dan *recall* sebesar 0.76. Namun, berdasarkan hasil eksperimen, performa model dapat ditingkatkan jika *Focal Loss* dikombinasikan dengan modifikasi distribusi label sehingga proporsi antar kelas menjadi lebih seimbang. Kombinasi tersebut menghasilkan performa terbaik dengan nilai *accuracy* sebesar 0.93, *precision* sebesar 0.90, *recall* sebesar 0.90, dan *F1-score* sebesar 0.90. Dengan demikian, *Focal Loss* cukup efektif dalam menangani ketidakseimbangan kelas. Namun, performa *Focal Loss* dapat lebih ditingkatkan dengan menerapkan strategi penyeimbangan distribusi label untuk memberikan hasil klasifikasi yang lebih optimal.
4. *User Interface* (UI) berbasis *website* telah berhasil dirancang dan diimplementasikan menggunakan *framework* Flask, SQLite, dan Tailwind CSS yang diintegrasikan dengan model IndoBERTweet-BiGRU. *Website* yang dikembangkan menyediakan fitur klasifikasi teks, visualisasi hasil analisis model, penjelasan alur kerja sistem, serta informasi mengenai penulis. Pengguna dapat membuat teks baru dan mengirim teks tersebut. Setelah menekan tombol kirim, sistem akan melakukan pendeteksian terhadap teks yang dimasukkan oleh pengguna. Jika teks tersebut mengandung *cyberbullying*, teks tersebut tidak akan terkirim dan pengguna akan

mendapatkan *pop-u*p yang menyatakan bahwa teks tersebut terindikasi *cyberbullying* dengan kategori tertentu. Sebaliknya, jika teks tersebut tidak mengandung *cyberbullying*, maka teks tersebut dapat dikirim dan disimpan oleh sistem. Dengan demikian, *website* dapat digunakan sebagai media interaktif untuk mempermudah pengguna dalam melakukan deteksi *cyberbullying* serta memahami alur dan performa model yang telah dikembangkan.

5.2. Saran Pengembangan

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa aspek yang dapat lebih dikembangkan untuk meningkatkan performa model maupun fungsionalitas sistem pada penelitian selanjutnya. Adapun saran pengembangan dari penelitian ini yaitu sebagai berikut.

1. Pada proses validasi hasil anotasi disarankan untuk melibatkan lebih dari satu anotator sehingga tingkat kesepakatan antar anotator (*inter-annotator agreement*) dapat dievaluasi secara lebih komprehensif dan menghasilkan kualitas anotasi yang lebih baik.
2. Proses pengumpulan data dapat dikembangkan dengan memperluas variasi kata kunci (*keyword*) yang digunakan pada tahap *scraping*. Hal tersebut diharapkan mampu menghasilkan *dataset* dengan variasi kata yang lebih beragam serta lebih merepresentasikan berbagai bentuk *cyberbullying* yang terdapat pada media sosial.
3. Penelitian selanjutnya disarankan untuk membandingkan performa beberapa model *Large Language Model* (LLM) pada proses anotasi data guna mengetahui model yang paling efektif dalam menghasilkan label yang akurat dan konsisten.
4. Pengembangan model deteksi *cyberbullying* dapat dilakukan dengan mempertimbangkan informasi selain teks, seperti pola interaksi pengguna, konteks percakapan, maupun metadata yang tersedia pada media sosial. Dengan demikian, sistem diharapkan mampu mengenali *cyberbullying* secara lebih komprehensif dan tidak hanya bergantung pada isi teks.