

BAB 5

SIMPULAN SARAN

5.1 Simpulan

Penelitian ini berhasil mengembangkan modul U-Tapis yang mampu mendeteksi kasus *harassment* berbasis penghinaan dan pencemaran nama baik secara otomatis menggunakan algoritma Naive Bayes Classifier. Modul ini dirancang dengan menggunakan dataset berbahasa Indonesia dan diuji dalam berbagai skenario pengujian untuk memenuhi tujuan penelitian. Sistem ini dibangun dengan framework Flask yang dapat memproses teks masukan melalui API berbasis *POST request*. Model yang digunakan telah dilatih menggunakan dataset yang terdiri dari 18.000 data untuk pelatihan dan 320 data untuk pengujian, yang mencakup tiga kategori: Netral, Hinaan, dan Pencemaran Nama Baik.

Evaluasi kinerja algoritma dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score pada berbagai ukuran dataset GPT (20, 50, 100, dan 150 sampel) dan 1177 dataset Internet dan Twitter. Hasil pengujian menunjukkan bahwa algoritma Naive Bayes memiliki akurasi berkisar antara 80% hingga 86%. Kinerja terbaik dicapai pada kategori Netral dan Pencemaran Nama Baik, dengan nilai *recall* hingga 98% pada dataset yang lebih besar. Namun, algoritma menghadapi tantangan dalam mengklasifikasikan teks pada kategori Hinaan, yang ditunjukkan dengan penurunan nilai *recall* pada kategori ini. Kesalahan klasifikasi yang terjadi antara kategori Hinaan dan Pencemaran Nama Baik menunjukkan perlunya analisis lebih mendalam terhadap pola linguistik yang membedakan kedua kategori tersebut.

Secara keseluruhan, modul U-Tapis mampu menjadi solusi awal untuk mendeteksi kasus *harassment* berbasis teks dalam bahasa Indonesia. Untuk peningkatan lebih lanjut, disarankan untuk memperluas dataset pelatihan dengan variasi konteks yang lebih beragam dan menggunakan metode pembelajaran yang lebih canggih, seperti *deep learning*, guna meningkatkan keandalan sistem dalam mendukung kebutuhan deteksi yang lebih kompleks.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan lebih lanjut modul U-Tapis agar lebih optimal dalam

mendeteksi kasus *harassment* berbasis penghinaan dan pencemaran nama baik, yaitu:

1. **Pengayaan Dataset:** Dataset yang digunakan dapat diperluas dengan menambahkan data dari berbagai sumber, seperti media sosial, forum diskusi, atau berita daring, sehingga pola linguistik yang lebih beragam dapat diakomodasi. Hal ini juga mencakup variasi bahasa informal atau campuran yang sering ditemukan dalam teks digital.
2. **Penggunaan Fitur Ekstraksi dengan Pretrained Model:** Untuk meningkatkan kemampuan model dalam menangkap pola dan konteks linguistik yang lebih kompleks, dapat digunakan fitur ekstraksi berbasis *pretrained model* seperti IndoBERT. Model ini mampu menghasilkan representasi teks yang lebih kaya dan kontekstual, sehingga dapat memperbaiki kualitas masukan yang digunakan oleh Naive Bayes. Dengan memanfaatkan representasi yang dihasilkan IndoBERT, algoritma seperti Naive Bayes dapat lebih efektif dalam mengklasifikasikan teks yang memiliki variasi bahasa atau konteks yang ambigu.
3. **Peningkatan Algoritma:** Penelitian lebih lanjut dapat mengeksplorasi algoritma berbasis *deep learning*, seperti LSTM, GRU, atau model berbasis *Transformer* (BERT, RoBERTa, atau GPT), untuk meningkatkan akurasi dan kemampuan model dalam menangkap pola linguistik yang kompleks, khususnya dalam kategori yang sulit dibedakan.