

BAB 5 SIMPULAN DAN SARAN

5.1 Simpulan

Penelitian ini berhasil mengembangkan sebuah modul deteksi hujatan berbahasa Indonesia, yang dibangun menggunakan algoritma *Support Vector Machine* (SVM). Modul ini dikembangkan dalam bentuk *pipeline* terintegrasi yang mencakup tahapan *text preprocessing*, ekstraksi fitur berbasis kombinasi *TF-IDF* tingkat kata dan karakter, serta fitur berbasis leksikon hujatan, yang kemudian diproses oleh model *LinearSVC* sebagai algoritma klasifikasinya. Pendekatan ini memungkinkan sistem untuk menangkap representasi teks yang lebih kaya, termasuk variasi ejaan kasar yang sering digunakan secara tidak eksplisit oleh pengguna media sosial.

Untuk mengoptimalkan performa model, penelitian ini menggunakan *Optuna* sebagai metode *hyperparameter tuning* berbasis *Bayesian Optimization*. Parameter-parameter yang dituning meliputi *max_features*, *min_df*, dan *ngram_range* pada *TF-IDF Word Vectorizer*, jumlah fitur dan rentang *ngram* pada *Char Vectorizer*, serta bobot kombinasi masing-masing fitur (*word_weight*, *char_weight*, dan *lexicon_weight*). Selain itu, parameter *C* dan *max_iter* pada *LinearSVC* juga turut dituning. Hasil *tuning* menunjukkan peningkatan akurasi dari 83% pada model awal menjadi 86% pada model hasil *tuning*, dengan peningkatan yang seimbang pada nilai *precision*, *recall*, dan *F1-score* untuk kedua kelas (hujatan dan non-hujatan).

Model hasil *tuning* kemudian diimplementasikan ke dalam sebuah aplikasi berbasis web dengan arsitektur *client-server*. Pada sisi *frontend*, digunakan *framework* *Next.js* untuk menyediakan antarmuka pengguna yang interaktif. Sementara itu, sisi *backend* dibangun menggunakan *Flask*, yang menyediakan *endpoint API* untuk mengakses model secara *real-time*. Implementasi ini menunjukkan bahwa model yang telah dikembangkan tidak hanya valid secara konseptual, tetapi juga dapat digunakan dalam konteks dunia nyata secara praktis dan efisien.

Namun demikian, model ini masih memiliki beberapa keterbatasan. Salah satunya adalah sensitivitas terhadap konteks semantik dan sarkasme, di mana sebuah kalimat mungkin mengandung kata kasar namun sebenarnya tidak

bermaksud menghina, atau sebaliknya. Selain itu, model juga dapat memberikan prediksi yang salah ketika teks mengandung kata-kata netral yang sering muncul dalam data pelatihan sebagai bagian dari hujatan. Oleh karena itu, pengembangan lebih lanjut diperlukan, misalnya dengan menambahkan pendekatan berbasis *contextual embedding* atau mempertimbangkan aspek pragmatik dalam analisis bahasa alami.

Meskipun secara umum sistem deteksi hujatan cenderung memiliki tingkat *false positive* (FP) yang tinggi akibat sensitivitas terhadap kata kasar atau bentuk variatif hujatan, hasil pada penelitian ini menunjukkan bahwa nilai FP (196) relatif seimbang dengan *false negative* (151), dan tidak terlalu dominan seperti yang umum terjadi pada sistem moderasi berbasis *keyword matching*. Salah satu faktor yang memungkinkan terjadinya hal ini adalah karena pendekatan representasi fitur yang digunakan dalam penelitian ini tidak hanya mengandalkan keberadaan kata kasar secara eksplisit, melainkan juga menggabungkan representasi konteks (TF-IDF) dan sinyal leksikal secara bersamaan. Dengan demikian, model memiliki kemampuan yang lebih baik dalam membedakan teks yang benar-benar hujatan dari teks netral yang kebetulan mengandung kata-kata kasar namun tidak berniat menghina. Kombinasi ini menghasilkan model yang tetap waspada terhadap hujatan, namun tidak terlalu agresif dalam mengklasifikasikan teks sebagai hujatan secara berlebihan.

Dengan demikian, penelitian ini telah menunjukkan bahwa kombinasi pendekatan klasik seperti SVM, teknik representasi fitur yang kaya, dan proses *tuning* yang sistematis dapat menghasilkan sistem deteksi hujatan yang cukup andal, sekaligus membuka ruang untuk pengembangan lebih lanjut di masa mendatang.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran untuk pengembangan lebih lanjut:

1. Penelitian selanjutnya disarankan untuk mengeksplorasi model berbasis *deep learning* seperti LSTM atau BERT untuk membandingkan performanya dengan model SVM, terutama dalam menangani konteks semantik yang kompleks.
2. Dataset yang digunakan dalam penelitian ini sebagian besar bersifat sintetik.

Oleh karena itu, pengumpulan data asli dari media sosial dengan anotasi oleh ahli bahasa dapat meningkatkan generalisasi dan keakuratan model pada kasus nyata.

3. Modul U-TAPIS dapat dikembangkan lebih lanjut untuk mendeteksi bentuk-bentuk *harassment* lain secara spesifik seperti ujaran kebencian berbasis SARA, atau pelecehan seksual, sehingga sistem memiliki cakupan yang lebih luas dalam mendeteksi ujaran kebencian secara komprehensif.
4. Fitur *feedback* dari pengguna terhadap hasil deteksi dapat ditambahkan ke dalam sistem untuk memperbaiki prediksi model di masa depan melalui pendekatan *active learning*.

Dengan adanya penelitian ini, diharapkan sistem yang dikembangkan dapat menjadi kontribusi nyata dalam menciptakan lingkungan digital yang lebih sehat dan bebas dari ujaran kebencian.

