

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan terkait analisis dan pemodelan data, dapat disimpulkan hal-hal berikut:

1. Pada hasil evaluasi model klasifikasi aspek, Stacking Ensemble tanpa Resampling menunjukkan performa tertinggi dibandingkan semua model individu lainnya, dengan F1-Score aspek Tampilan 99.4%, Autentikasi 99.3%, dan Transaksi 99%. Meskipun penerapan SMOTE tetap menghasilkan performa yang tinggi, performa terbaik tetap ada pada Stacking Ensemble tanpa SMOTE. Hal ini menunjukkan bahwa model sudah mampu menangani ketidakseimbangan kelas secara efektif tanpa perlu resampling, sehingga penggunaan SMOTE tidak memberikan peningkatan yang signifikan terhadap hasil klasifikasi aspek.
2. Pada hasil evaluasi model klasifikasi sentimen tiap aspek, Stacking Ensemble tanpa Resampling juga menunjukkan performa tertinggi dibandingkan semua model individu lainnya. Skor F1-score untuk sentimen aspek Tampilan 82.2%, Sentimen Autentikasi 87.8%, dan Sentimen Transaksi: 92.4%. Setelah penerapan SMOTE, performa model tetap tinggi tetapi tidak menunjukkan peningkatan yang berarti. Hal ini membuktikan bahwa Stacking Ensemble sudah cukup andal dalam menangani ketidakseimbangan data secara alami, sehingga metode resampling seperti SMOTE tidak memberikan dampak yang besar terhadap performanya melainkan lebih berdampak pada algoritma individu.
3. Penerapan *Local Interpretable Model-agnostic Explanations* (LIME) dalam analisis ulasan dengan model *stacking ensemble* terbukti mampu memberikan penjelasan yang transparan mengenai pengambilan keputusan model. Hasil interpretasi menunjukkan bahwa LIME dapat mengidentifikasi kata-kata yang paling berkontribusi terhadap klasifikasi aspek dan sentimen dalam ulasan pengguna, baik dalam kategori *positif*, *negatif*, maupun berbagai aspek seperti *tampilan*, *otentikasi*, dan *transaksi*. Dengan adanya visualisasi berupa

highlight pada kata-kata penting dan bobot kontribusi masing-masing kata, pengguna dapat memahami bagaimana model menentukan hasil prediksi. LIME juga dapat digunakan untuk mengidentifikasi kesalahan dalam klasifikasi. Oleh karena itu, LIME dapat menjadi alat yang efektif dalam meningkatkan transparansi dan interpretabilitas model *stacking ensemble*, terutama dalam tugas analisis sentimen berbasis aspek (*Aspect-Based Sentiment Analysis*).

4. Pengembangan website analisis aspek sentimen dilakukan dengan beberapa tahapan. Pertama, model terbaik yang telah dilatih, yaitu *stacking ensemble*, diekspor ke dalam format *.pkl*. Model ini sudah mencakup pipeline preprocessing, vectorizer, dan model itu sendiri. Setelah itu, backend dikembangkan menggunakan framework Flask. Backend ini berfungsi untuk memuat file model *.pkl*, memproses input teks dari pengguna, melakukan prediksi aspek dan sentimen, lalu mengirimkan hasilnya melalui API. Frontend dibuat menggunakan HTML, CSS, dan JavaScript dalam kerangka Flask, dan dihubungkan dengan backend melalui API. Pada frontend, pengguna dapat memasukkan atau mengunggah ulasan. Hasil analisis aspek dan sentimen kemudian ditampilkan dalam bentuk tabel atau visualisasi, sehingga memberikan wawasan yang bermanfaat dalam pengambilan keputusan berbasis data.

5.2. Saran Pengembangan

Berikut ini adalah beberapa usulan yang dapat mendukung pengembangan studi lebih lanjut:

1. Bagi peneliti selanjutnya dapat mempertimbangkan dalam melakukan normalisasi teks lebih mendalam untuk menyamakan variasi kata dengan makna yang serupa, sehingga model lebih konsisten dalam memahami input dan meningkatkan probabilitas model dalam klasifikasi.
2. Menerapkan contextual embedding seperti BERT dan model BERT untuk meningkatkan pemahaman konteks antar kata dan mengurangi ketergantungan pada kata kunci tertentu, serta menggunakan pendekatan binary relevance untuk memungkinkan prediksi sentimen dilakukan secara independen untuk masing-masing aspek, sehingga kesalahan pengaitan sentimen antar aspek dapat diminimalkan.

Pengembang aplikasi Wondr by BNI dapat mempertimbangkan beberapa saran berikut:

1. Bagi pengembang aplikasi disarankan untuk mempertahankan kualitas tampilan antarmuka yang sudah mendapat apresiasi positif dari pengguna, serta memfokuskan perbaikan pada aspek autentikasi dan transaksi. Tingginya keluhan pada proses login dan verifikasi menunjukkan perlunya peningkatan keamanan dan kemudahan akses, sementara banyaknya ulasan negatif pada fitur transaksi mengindikasikan masih adanya masalah teknis yang perlu segera diatasi untuk meningkatkan kenyamanan dan kepercayaan pengguna.
2. Memanfaatkan website prediksi aspek dan sentimen untuk analisis otomatis. Website ini dapat membantu tim pengembangan aplikasi Wondr by BNI dalam memahami masukan pengguna secara lebih efisien. Dengan mengelompokkan ulasan ke dalam aspek seperti tampilan, autentikasi, dan transaksi, serta mengidentifikasi sentimen (positif atau negatif), tim dapat mengetahui bagian aplikasi mana yang paling diapresiasi atau justru paling banyak dikeluhkan. Hal ini sangat bermanfaat untuk menentukan prioritas pengembangan, mengevaluasi hasil perubahan desain, dan membuat keputusan berbasis data tanpa harus membaca ribuan ulasan secara manual.

Halaman ini sengaja dikosongkan