

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian dan analisis komparasi yang telah dilakukan, pemanfaatan dataset yang berisi fitur struktural URL yang telah terencode terbukti sangat efektif dan efisien dalam mendeteksi ancaman website phishing. Kontribusi utama pada penelitian ini difokuskan pada tahap pemilihan dataset, pra-pemrosesan (preprocessing), serta evaluasi komparatif model klasifikasi, bukan pada ekstraksi fitur dari raw URL secara mandiri. Penggunaan representasi fitur yang mencakup pola leksikal, tingkat keamanan protokol, dan metrik reputasi domain ini memungkinkan sistem klasifikasi untuk mencapai tingkat akurasi yang tinggi, yakni secara merata berada di atas 96%. Secara teoritis, pendekatan berbasis struktur URL memiliki potensi keunggulan efisiensi komputasional dibandingkan pendekatan berbasis konten, karena tidak memerlukan pengunduhan dan analisis mendalam terhadap elemen halaman web.

Dari ketiga algoritma *ensemble learning* yang dievaluasi, LightGBM berhasil menempati posisi sebagai model dengan kinerja prediktif terbaik dan paling seimbang secara keseluruhan. Melalui pengujian *Stratified 5-Fold Cross Validation*, LightGBM mendominasi dengan tingkat rata-rata akurasi sebesar 97,01%, *Recall* 97,85%, *F1-Score* 97,33%, dan nilai *ROC-AUC* mencapai 99,66%. Keunggulan performa ini secara arsitektural didorong oleh mekanisme *leaf-wise tree growth* yang memungkinkannya mengekstraksi pola anomali URL secara tajam sekaligus menekan angka *False Positive* atau kesalahan pemblokiran seminimal mungkin.

Sementara itu, algoritma Random Forest mencatatkan diri sebagai model dengan tingkat presisi (*Precision*) paling tinggi dibandingkan model lainnya, yakni mencapai rata-rata 97,33% pada pengujian *cross-validation*. Mekanisme agregasi paralel (*majority voting*) pada arsitektur Random Forest terbukti mampu membangun sistem filter analitis yang sangat ketat. Karakteristik perlindungan maksimal ini menjadikannya pilihan algoritma yang sangat direkomendasikan

apabila prioritas utama dari implementasi sistem keamanan di dunia nyata adalah untuk meminimalisir kesalahan pemblokiran pada *website* publik yang sebenarnya sah (menghindari alarm palsu).

Secara keseluruhan, penerapan metode evaluasi ganda menggunakan *Hold-out Split* dengan rasio 80:20 dan *Stratified 5-Fold Cross Validation* berhasil mengonfirmasi bahwa ketiga algoritma (XGBoost, LightGBM, dan Random Forest) memiliki stabilitas komputasi yang sangat baik. Berdasarkan analisis visual pada grafik *learning curve*, penambahan jumlah observasi data latih terbukti berhasil membantu ketiga model untuk mengonvergensi tingkat galat (*error*) secara mulus. Hal ini memastikan bahwa model-model yang dilatih terlepas dari jebakan *overfitting* serta memiliki ketangguhan dalam menggeneralisasi dan memprediksi data URL baru (*unseen data*) saat diimplementasikan ke dalam sistem.

5.2 Saran

Berdasarkan kesimpulan yang telah diuraikan serta batasan-batasan yang ada selama pelaksanaan penelitian ini, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan dan penyempurnaan penelitian selanjutnya:

1. Eksplorasi Kombinasi Fitur Tambahan: Penelitian ini memiliki ruang lingkup yang secara spesifik dibatasi hanya pada penggunaan fitur struktur URL tanpa melibatkan analisis terhadap konten antarmuka maupun fitur berbasis jaringan (*network-based features*). Untuk meningkatkan ketangguhan deteksi terhadap teknik *phishing* yang terus berkembang dan semakin manipulatif, penelitian selanjutnya disarankan untuk menggabungkan fitur leksikal URL dengan fitur berbasis konten (seperti analisis elemen HTML, teks halaman, atau *script* halaman web).
2. Pengembangan Implementasi Deteksi *Real-Time*: Pada penelitian ini, sistem tidak dieksplorasi hingga tahap implementasi secara *real-time* atau diintegrasikan langsung ke dalam aplikasi keamanan web. Meskipun model telah berhasil diimplementasikan ke dalam bentuk *prototype* aplikasi berbasis *web* menggunakan Streamlit, pemanfaatan praktisnya akan jauh lebih optimal apabila model pendeteksi (khususnya LightGBM)

dikembangkan menjadi sebuah ekstensi peramban (*browser extension* atau *add-on*). Integrasi ini akan memungkinkan sistem bekerja memindai anomali URL secara otomatis di latar belakang dan memberikan peringatan *real-time* kepada pengguna sebelum mereka memuat halaman situs berbahaya.

3. Eksplorasi Pendekatan *Deep Learning*: Meskipun algoritma *ensemble learning* tradisional (XGBoost, LightGBM, dan Random Forest) telah terbukti mampu memberikan akurasi di atas 96% , penelitian berikutnya dapat mulai mengeksplorasi pemanfaatan arsitektur *Deep Learning* seperti *Convolutional Neural Networks (CNN)* atau *Long Short-Term Memory (LSTM)*. Sebagai catatan teknis, model *deep learning* tersebut memiliki keunggulan komputasional dalam memproses input berupa deret karakter mentah (*raw character sequence*) dari URL untuk mengekstraksi fitur sekuensial secara otomatis. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan *dataset* yang memuat *raw URL*, bukan *dataset* dengan representasi fitur numerik diskrit seperti yang diterapkan pada penelitian ini, guna mereduksi ketergantungan pada tahapan rekayasa fitur (*feature engineering*) manual.
4. Pembaruan dan Peningkatan *Volume Dataset*: Mengingat arsitektur, teknik penyembunyian, dan pola penamaan domain pada ancaman *website phishing* bervariasi dan berubah dengan sangat cepat seiring berjalannya waktu, disarankan agar studi selanjutnya menggunakan *dataset* dengan cakupan sampel yang lebih masif dan mutakhir. Pemanfaatan sumber data yang diperbarui secara dinamis akan sangat membantu dalam melatih model agar lebih adaptif dan kebal terhadap tren serangan siber generasi terbaru.