BAB V
PENUTUP

5.1 Kesimpulan

Penelitian ini bertujuan sebagai bentuk optimalisasi deteksi *URL phishing* dengan menggunakan metode *feature subset ranking* dengan menggunakan fitur-fitur dengan tingkat kepentingan paling tinggi dan model pembelajaran mesin menggunakan *lightweight models*. Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan beberapa hal berikut:

1. Penerapan metode *feature subset ranking* menggunakan *Random Forest Feature Importance* secara signifikan meningkatkan efisiensi proses pelatihan model. Reduksi jumlah fitur dari 89 menjadi 20 fitur (berkurang 77.53%), yang berdampak pada waktu pelatihan model yang lebih singkat hingga rata- rata 81.85% dan struktur model yang lebih ringan. Eksperimen lanjutan juga menunjukan bawasanya penggunaan hanya 5 atau 10 fitur tetap menghasilkan performa kompetitif, meskipun terjadi penurunan akurasi yang wajar. Misalnya, dengan hanya 10 fitur, model *CatBoost* masih mencatat akurasi 95.23% dan *F1-score* 95.16% yang tinggi, menjadikannya tetap efektif dengan struktur model yang jauh lebih ringan.
2. Model *CatBoost* terbukti paling optimal dalam hal kesesimbangan akurasi, efisiensi waktu pelatihan, dan kecepatan inferensi. untuk mendeteksi phishing berbasis *URL* dari segi akurasi dan kecepatan inferensi. Dengan akurasi 96.54% dan *F1-Score* 96.49% menggunakan 20 fitur, serta waktu inferensi hanya 3.39 milidetik, model ini sangat cocok untuk diterapkan pada sistem deteksi *phishing real-time*. *CatBoost* juga menunjukan *precision* dan *recall* yang seimbang, membuktikan kemampuannya dalam menjaga tingkat deteksi yang tinggi tanpa mengorbankan performa sistem.

5.2 Saran

Terdapat beberapa saran yang dapat diberikan untuk pengembangan lebih lanjut terkait penelitian ini, antara lain:

1. Penambahan Jenis Data: Untuk meningkatkan generalisasi model, disarankan untuk menggunakan dataset yang lebih beragam, termasuk data *real-time* ataupun data dari berbagai sumber geografis.
2. Eksplorasi Teknik seleksi fitur Lain: Selain *random forest*, teknik seperti *Recursive Feature Elimination* (RFE) atau *mutual information* dapat dieksplorasi untuk membandingkan efektifitas seleksi fitur.
3. Pengembangan Antarmuka GUI/WEB: Aplikasi CLI dapat dikembangkan lebih lanjut menjadi antarmuka berbasis web supaya mempermudah dan lebih ramah pengguna dan juga dapat digunakan oleh kalangan non-teknis.
4. Penanganan *False Positive*: Perlu dilakukan penelitian lanjutan untuk menurunkan tingkat *false positive*, terutama pada model *Random Forest* yang cenderung memberikan hasil positif berlebih.
5. Integrasi Keamanan Tambahan: Sistem deteksi ini dapat digabungkan dengan teknik keamanan lainnya seperti analisis konten *HTML* atau sistem deteksi berbasis DNS untuk memberikan perlindungan lebih menyeluruh.