BAB V
KESIMPULAN

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan algoritma *CatBoost* untuk mendeteksi *ransomware* berdasarkan karakteristik file eksekusi. Model yang dikembangkan menunjukkan performa sangat tinggi dengan nilai ROC AUC 0.9972, serta *precision, recall,* dan *F1-score* sebesar 0.9902, yang mencerminkan prediksi yang akurat dan seimbang. Keunggulan *CatBoost* terletak pada kemampuannya menangani fitur kategorikal tanpa banyak *preprocessing*, mengelola data dengan banyak fitur, dan menangani *missing values*.

Proses implementasi meliputi pengumpulan dataset yang mencakup karakteristik file *ransomware* dan *benign*, ekstraksi fitur seperti tipe arsitektur mesin, ukuran data debug, ukuran tabel ekspor, karakteristik dari file dll, dan lainnya. Model dilatih dengan membagi dataset menjadi data latih dan uji, menggunakan parameter *class weight* untuk menangani ketidakseimbangan kelas. Kemudian evaluasi dilakukan dengan metrik akurasi, *precision, recall, F1-score, Logloss, Confussion matrix,* dan ROC AUC untuk memastikan performa optimal.

Model yang telah dilatih diintegrasikan ke *website* berbasis *Flask*, memungkinkan pengguna mengunggah file untuk dianalisis secara otomatis dan mendapatkan prediksi apakah file tersebut *ransomware* atau bukan. Optimasi lebih lanjut dilakukan melalui *tuning hyperparameter* dan pengujian dengan dataset baru untuk meningkatkan akurasi dan generalisasi.

Karakteristik file eksekusi yang paling berpengaruh dalam meningkatkan akurasi deteksi *ransomware* meliputi atribut PE (*Portable Executable*), seperti tipe arsitektur mesin, ukuran data debug, ukuran tabel ekspor, karakteristik dari file dll, alamat bitcoin, dan lain-

lainnya. Dengan mempertimbangkan fitur-fitur ini, model dapat mengenali pola yang membedakan *ransomware* dari file *benign*.

Meskipun dataset penelitian ini tidak seimbang antara jumlah file *benign* yang lebih banyak dibandingkan dengan file yang terinfeksi *ransomware* tetapi *CatBoost* mampu mengatasinya dengan baik berkat teknik *boosting* yang digunakan serta pemanfaatan metode seperti parameter *class weight* dalam tahap *preprocessing* data. Hasil evaluasi menunjukkan bahwa model tetap dapat menghasilkan prediksi yang akurat tanpa mengalami *overfitting* atau bias terhadap kelas mayoritas.

5.2 Saran

1. Penelitian selanjutnya dapat mengintegrasikan teknik interpretabilitas model, seperti SHAP (*Shapley Additive Explanations*), untuk memahami lebih dalam bagaimana setiap fitur berkontribusi terhadap keputusan klasifikasi. Penggunaan *ensemble learning* dengan mengombinasikan *CatBoost* dengan model lain seperti *Random Forest* atau *XGBoost* dapat dieksplorasi untuk meningkatkan performa sistem deteksi.
2. Penelitian lanjutan dapat mengeksplorasi karakteristik tambahan seperti *API calls* yang dilakukan oleh file eksekusi atau analisis *behavior-based detection* untuk meningkatkan deteksi terhadap varian *ransomware* terbaru. Pengujian model pada dataset yang lebih besar dan lebih bervariasi dapat memberikan wawasan lebih lanjut mengenai fitur mana yang paling konsisten dalam membedakan *ransomware* dari file *benign*.
3. Menggunakan teknik sampling yang lebih canggih, seperti *Adaptive Synthetic Sampling* (ADASYN), dapat meningkatkan keseimbangan dataset tanpa menambahkan *noise* yang signifikan. Eksperimen dengan *threshold tuning* untuk mengoptimalkan *trade-off* antara *precision* dan *recall*, terutama dalam *kondisi real-world* di mana *false positives* dapat berdampak pada operasional sistem keamanan.