

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	ii
HALAMAN PERSETUJUAN.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERNYATAAN KEASLIAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR ISTILAH	xiv
DAFTAR LAMPIRAN.....	xvii
ABSTRAK.....	xviii
<i>ABSTRACT</i>	xix
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Rumusan Masalah.....	6
C. Batasan Masalah	6
D. Tujuan Penelitian	7
E. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA.....	9
A. Landasan Teori.....	9
B. Penelitian Sebelumnya.....	43
BAB III METODE PENELITIAN.....	50
A. Tempat dan Waktu Penelitian.....	50
B. Metode Pengumpulan Data.....	50
C. Alat dan Bahan Penelitian.....	54

D. Konsep Penelitian	58
BAB IV HASIL DAN PEMBAHASAN	71
A. Pengumpulan Data	71
B. Praproses Data	72
C. Pembagian Data	76
D. Seleksi Fitur Mutual Information.....	79
E. Arsitektur Model TabNet.....	84
F. Skenario dan Desain Eksperimen	87
G. Metrik Evaluasi.....	91
H. Pembahasan.....	98
BAB V PENUTUP.....	107
A. Kesimpulan	107
B. Saran	107
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR TABEL

Tabel 2.1. Struktur Confusion Matrix Klasifikasi Biner.....	36
Tabel 2.2. Penelitian Sebelumnya.....	48
Tabel 3.1. Jadwal Pelaksanaan Penelitian.....	50
Tabel 3.2. Karakteristik Dataset MH-100K.....	52
Tabel 3.3. Spesifikasi Perangkat Keras Penelitian.....	55
Tabel 3.4. Spesifikasi Perangkat Lunak Penelitian.....	56
Tabel 3.5. Ringkasan Hasil Praproses Data	61
Tabel 3.6. Hasil Pembagian Dataset	63
Tabel 3.7. Skenario Seleksi Fitur Berdasarkan Threshold MI.....	65
Tabel 3.8. Konfigurasi Hyperparameter Model TabNet.....	67
Tabel 4.1. Hasil Pengumpulan dan Pemuatan Dataset MH-100K.....	71
Tabel 4.2. Distribusi Kelas Target Setelah Deduplikasi	74
Tabel 4.3. Ringkasan Hasil Tahapan Praproses Data	76
Tabel 4.4. Hasil Pembagian Dataset per Partisi	78
Tabel 4.5. Statistik Deskriptif Skor Mutual Information	80
Tabel 4.6. Lima Fitur Tertinggi Berdasarkan Skor Mutual Information	81
Tabel 4.7. Hasil Seleksi Fitur Berdasarkan Empat Nilai Threshold	83
Tabel 4.8. Perbandingan Hasil Pelatihan Model TabNet.....	89
Tabel 4.9. Confusion Matrix Model Baseline TabNet.....	92
Tabel 4.10. Confusion Matrix Model Final TabNet + MI-Q3	94
Tabel 4.11. Perbandingan Metrik Evaluasi.....	95
Tabel 4.12. Ringkasan Dampak Penerapan Seleksi Fitur MI-Q3	104

DAFTAR GAMBAR

Gambar 2.1. Arsitektur Android	11
Gambar 2.2. Arsitektur TabNet.....	26
Gambar 3.1. Flowchart Alur Penelitian	59
Gambar 4.1. Distribusi dan Proporsi Kelas Target	75
Gambar 4.2. Distribusi Kelas pada Data Latih, Data Validasi, dan Data Uji	78
Gambar 4.3. Top 20 Fitur Berdasarkan Skor Mutual Information	82
Gambar 4.4. Confusion Matrix Model Baseline TabNet	93
Gambar 4.5. Confusion Matrix Model Final TabNet + MI-Q3	94
Gambar 4.6. Kurva Precision-Recall Model Baseline dan Model Final MI-Q3...	97
Gambar 4.7. Top 20 Fitur Terpenting Berdasarkan Attention Importance TabNet	101

DAFTAR ISTILAH

Accuracy : Metrik evaluasi yang mengukur proporsi prediksi benar dari keseluruhan sampel yang diuji.

Android Package Kit (APK) : Format berkas arsip yang memuat seluruh komponen aplikasi Android, termasuk kode program, sumber daya, dan berkas manifes.

API Call : Pemanggilan terhadap fungsi Application Programming Interface yang terdapat dalam kode program suatu aplikasi.

Attention Mask : Bobot perhatian yang dihasilkan oleh mekanisme attention untuk menentukan tingkat kepentingan suatu fitur dalam proses pengambilan keputusan model.

Attentive Transformer : Komponen arsitektur TabNet yang menghasilkan attention mask untuk menentukan fitur yang akan diproses pada setiap langkah.

Area Under the Precision-Recall Curve (AUPRC) : Metrik evaluasi yang mengukur luas area di bawah kurva Precision-Recall, digunakan untuk mengevaluasi performa model pada kondisi data tidak seimbang.

Backpropagation : Algoritma pelatihan jaringan saraf tiruan yang menghitung gradien kesalahan untuk memperbarui bobot model secara mundur dari lapisan keluaran ke lapisan masukan.

Baseline : Model atau kondisi acuan yang digunakan sebagai pembanding untuk mengevaluasi peningkatan performa metode yang diusulkan.

Batch Size : Jumlah sampel data yang diproses secara bersamaan dalam satu kali iterasi pelatihan model.

Benign : Klasifikasi aplikasi yang tidak mengandung unsur berbahaya atau bersifat aman bagi pengguna.

Class Imbalance : Kondisi ketidakseimbangan jumlah sampel antar kelas dalam suatu dataset klasifikasi.

Confusion Matrix : Tabel yang menyajikan ringkasan hasil perbandingan antara prediksi model dan label kelas yang sesungguhnya.

Deep Learning : Subkategori machine learning yang menggunakan jaringan saraf tiruan dengan banyak lapisan tersembunyi untuk mempelajari representasi data secara hierarkis.

Early Stopping : Teknik penghentian proses pelatihan model secara otomatis ketika tidak terjadi peningkatan performa pada data validasi selama sejumlah epoch tertentu.

Entropi Shannon : Ukuran ketidakpastian rata-rata atau kandungan informasi dari suatu variabel acak diskret.

Epoch : Satu siklus lengkap proses pelatihan model yang melibatkan seluruh data latih.

F1-Score : Metrik evaluasi yang merupakan rata-rata harmonik dari precision dan recall.

False Negative (FN) : Jumlah sampel positif (Malware) yang salah diklasifikasikan sebagai negatif (Benign).

False Positive (FP) : Jumlah sampel negatif (Benign) yang salah diklasifikasikan sebagai positif (Malware).

Feature Importance : Skor yang menunjukkan tingkat kontribusi relatif suatu fitur terhadap prediksi model.

Feature Selection (Seleksi Fitur) : Proses pemilihan subset fitur paling relevan dari keseluruhan fitur yang tersedia dalam dataset.

Feature Transformer : Komponen arsitektur TabNet yang memproses fitur-fitur terseleksi untuk menghasilkan representasi pada setiap langkah pemrosesan.

Filter Method : Kategori metode seleksi fitur yang mengevaluasi relevansi fitur secara independen dari algoritma pembelajaran yang digunakan.

Hyperparameter : Parameter konfigurasi model yang ditentukan sebelum proses pelatihan dan tidak dipelajari langsung dari data.

Intent : Mekanisme pesan antar komponen aplikasi Android yang mendefinisikan aksi dan data yang dapat direspons oleh suatu komponen.

Malware : Singkatan dari malicious software, yaitu perangkat lunak yang dirancang untuk menyusup atau merusak sistem tanpa izin pengguna.

Mutual Information (MI) : Ukuran ketergantungan statistik antara dua variabel acak yang mengukur jumlah informasi yang dibagi secara bersama.

Overfitting : Kondisi suatu model yang terlalu menyesuaikan diri terhadap data latih sehingga performanya menurun pada data baru.

Permission (Izin) : Deklarasi akses yang dibutuhkan suatu aplikasi Android untuk mengakses sumber daya sistem tertentu.

Precision : Metrik evaluasi yang mengukur proporsi prediksi positif yang benar-benar merupakan kelas positif.

Random Seed : Nilai awal yang digunakan untuk menginisialisasi proses pengacakan agar hasil eksperimen dapat direproduksi.

Recall : Metrik evaluasi yang mengukur proporsi sampel positif aktual yang berhasil terdeteksi oleh model.

Sequential Attention : Mekanisme attention bertahap pada arsitektur TabNet yang memilih subset fitur berbeda di setiap langkah pemrosesan.

Sparsemax : Fungsi aktivasi yang menghasilkan distribusi probabilitas jarang (sparse), digunakan dalam mekanisme attention TabNet.

Stratified Split : Teknik pembagian dataset yang mempertahankan proporsi kelas yang konsisten pada setiap partisi data.

TabNet : Arsitektur deep learning yang dirancang khusus untuk data tabular dengan memanfaatkan mekanisme sequential attention.

Threshold : Nilai ambang batas yang digunakan sebagai acuan dalam pengambilan keputusan, baik dalam seleksi fitur maupun klasifikasi.

True Negative (TN) : Jumlah sampel negatif (Benign) yang diprediksi dengan benar sebagai negatif.

True Positive (TP) : Jumlah sampel positif (Malware) yang diprediksi dengan benar sebagai positif.

DAFTAR LAMPIRAN

Lampiran 1. Kartu Bimbingan Dosen Pembimbing 1

Lampiran 2. Listing Program — Praproses Data dan Seleksi Fitur MI

Lampiran 3. Listing Program — Pelatihan Model TabNet Baseline

Lampiran 4. Listing Program — Pelatihan Model TabNet dengan Seleksi Fitur MI-Q3

