

**PENDETEKSIAN MALWARE ANDROID DENGAN MEMANFAATKAN
FITUR PERMISSIONS MENGGUNAKAN ALGORITMA EXTREME
GRADIENT BOOSTING (XGBOOST)**

Tugas Akhir

Diajukan sebagai salah satu syarat memperoleh gelar Sarjana
Program Studi S-1 Teknologi Informasi



Disusun oleh:

NAFISA ALFI SA'DIYA

212104006

**PROGRAM STUDI S-1 TEKNOLOGI INFORMASI
FAKULTAS TEKNIK & TEKNOLOGI INFORMASI
UNIVERSITAS JENDERAL ACHMAD YANI YOGYAKARTA
2025**

HALAMAN PENGESAHAN

TUGAS AKHIR
PENDETEKSIAN MALWARE ANDROID DENGAN MEMANFAATKAN
FITUR PERMISSIONS MENGGUNAKAN ALGORITMA EXTREME
GRADIENT BOOSTING (XGBOOST)

Diajukan oleh:

NAFISA ALFI SA'DIYA
212104006

Telah dipertahankan di depan dewan penguji dan dinyatakan sah
sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta

Tanggal: 11 Juli 2025

Mengesahkan:

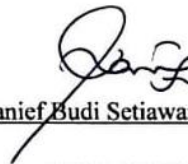
Pembimbing



Arief Ikhwana Wicaksono., S.Kom., M.Cs.

NIDN: 0512128401

Penguji



Chanief Budi Setiawan, S.T., M.Eng.

NIDN: 0514068101

Ketua Program Studi S-1 Teknologi Informasi
Fakultas Teknik & Teknologi Informasi
Universitas Jenderal Achmad Yani Yogyakarta



Rama Sahtyawan, S.T., M.Cs.
NPP: 2019.13.00150

PERNYATAAN

Saya yang bertanda tangan di bawah ini, adalah mahasiswa Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta,

Nama : Nafisa Alfi Sa'diya
NPM : 212104006
Program Studi : S-1 Teknologi Informasi
Judul Tugas Akhir : Pendeteksian Malware Android dengan Memanfaatkan Fitur Permissions Menggunakan Algoritma Extreme Gradient Boosting (XGBoost)

Menyatakan bahwa hasil penelitian dengan judul tersebut di atas adalah asli karya saya sendiri dan bukan hasil plagiarisme. Semua referensi dan sumber terkait yang dikutip dalam karya ilmiah ini telah ditulis sesuai kaidah penulisan ilmiah yang berlaku. Dengan ini, saya menyatakan untuk menyerahkan hak cipta penelitian kepada Universitas Jenderal Achmad Yani Yogyakarta guna kepentingan ilmiah.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya tanpa ada paksaan dari pihak mana pun. Apabila terdapat kekeliruan atau ditemukan adanya pelanggaran akademik di kemudian hari, maka saya bersedia menerima konsekuensi yang berlaku sesuai ketentuan akademik.

Yogyakarta, 20 Juni 2025



Nafisa Alfi Sa'diya

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul: “Pendeteksian Malware Android dengan Memanfaatkan Fitur Permissions Menggunakan Algoritma Extreme Gradient Boosting (XGBoost)”. Penyusunan laporan ini merupakan salah satu persyaratan untuk menyelesaikan studi di Program Studi S-1 Teknologi Informagambarsi Fakultas Teknik & Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta. Laporan ini dapat diselesaikan atas bimbingan, arahan, dan bantuan dari berbagai pihak. Pada kesempatan ini penulis dengan rendah hati mengucapkan terima kasih dengan setulus-tulusnya kepada:

1. Bapak Aris Wahyu Murdiyanto, S.Kom., M.Cs. selaku Dekan Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
2. Bapak Rama Sahtyawan, S.T., M.Cs. selaku Ketua Program Studi S-1 Teknologi Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
3. Bapak Arief Ikhwan Wicaksono., S.Kom., M.Cs. selaku Dosen Pembimbing Tugas Akhir;
4. Para dosen yang telah memberikan banyak bekal ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta;
5. Ayah, ibu, kakak dan adik-adik saya yang telah memberikan dukungan semangat serta doa restu kepada saya, sehingga dapat menyelesaikan studi saya;
6. Sahabat-sahabat terbaik, Kartikasari, Ria, dan Alfun atas dukungan emosional, semangat, dan tawa yang selalu menyertai;

7. Mark Lee, sebagai *bias* sekaligus *favorite person* saya, yang menjadi sumber semangat dan penghibur melalui pesan-pesan positif dan inspiratif. Dari Mark, saya belajar arti kerja keras, ketekunan, serta bagaimana untuk tidak menyerah meskipun dalam keadaan sulit.
8. Terakhir, terima kasih banyak untuk diri saya sendiri! Terima kasih atas keberanian untuk tetap bertahan, menyelesaikan, dan terus berproses walaupun di tengah berbagai kesulitan dan tekanan. Terima kasih telah terus berjalan meski tidak mudah, tetap bekerja walau disertai air mata, dan tetap hidup meskipun banyak hal membuat ingin menyerah.

Penulis menyadari bahwa laporan tugas akhir ini masih jauh dari kata sempurna. Maka dari itu dengan segala kerendahan hati penulis sangat menghargai adanya kritik dan saran yang membangun dari semua pihak yang bersedia meluangkan waktu untuk membaca laporan tugas akhir ini.

Yogyakarta, 20 Juni 2025



Nafisa Alfi Sa'diya

DAFTAR ISI

Halaman Pengesahan.....	ii
Pernyataan	iii
Kata Pengantar	iv
Daftar Isi	vi
Daftar Tabel.....	ix
Daftar Gambar	x
Daftar Lampiran	xi
Daftar Singkatan	xii
Intisari	xiii
Abstract.....	xiv
BAB 1 Pendahuluan.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Pertanyaan Penelitian	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Hasil Penelitian	5
1.6 Batasan Penelitian	5
BAB 2 Tinjauan Pustaka dan Landasan Teori	6
2.1 Tinjauan Pustaka	6
2.2 Landasan Teori.....	13
2.2.1 Malware.....	13
2.2.2 Android.....	13
2.2.3 <i>App Permissions</i>	14
2.2.4 Machine Learning	14
2.2.5 <i>Imbalanced Class</i>	15
2.2.6 Extreme Gradient Boosting (XGBoost)	15
2.2.7 <i>Hyperparameter Tuning</i>	18
2.2.8 Evaluasi Model Machine Learning	19

2.2.9	<i>Threshold Tuning</i>	23
2.2.10	<i>Synthetic Minority Oversampling Technique (SMOTE)</i>	24
BAB 3 Metode Penelitian		25
3.1	Tahapan Penelitian	25
3.1.1	Dataset Penelitian	27
3.1.2	<i>Preprocessing Data</i>	28
3.1.3	<i>Data Splitting</i>	28
3.1.4	<i>Hyperparameter Tuning</i>	29
3.1.5	Pelatihan Model XGBoost	29
3.1.6	Evaluasi Model.....	30
3.1.7	Analisis Hasil	32
3.1.8	Eksperimen Subset Data.....	34
3.1.9	<i>Threshold Tuning</i>	34
3.1.10	Penerapan SMOTE.....	35
3.2	Bahan Penelitian.....	36
3.3	Alat Penelitian.....	37
3.4	Jalan Penelitian.....	37
BAB 4 Hasil Penelitian		39
4.1	Ringkasan Hasil Penelitian	39
4.2	Pembahasan Hasil Pemeodelan.....	40
4.2.1	<i>Preprocessing</i>	42
4.2.2	Eksplorasi Distribusi Kelas	45
4.2.3	<i>Baseline Model</i>	46
4.2.4	Perhitungan <i>scale_pos_weight</i>	47
4.2.5	Split Data.....	49
4.2.6	Pelatihan Model XGBoost Awal.....	50
4.2.7	Evaluasi Awal Model	52
4.2.8	<i>Hyperparameter Tuning</i>	54
4.2.9	Evaluasi Model Terbaik	57
4.2.10	Evaluasi Lanjutan: ROC, AUC, dan Nilai TP-FP-FN-TN.....	58
4.2.11	Pengambilan Subset Data.....	61

4.2.12 Rekap dan Visualisasi Evaluasi.....	69
4.2.13 Threshold Tuning	73
4.2.14 Penerapan SMOTE.....	76
4.2.15 Perbandingan Model	80
4.2.16 Decision Tree dan Karakteristik Malware Berdasarkan Permissions	81
4.3 Analisis Hasil	88
BAB 5 Kesimpulan Dan Saran.....	90
5.1 Kesimpulan	90
5.2 Saran.....	91
Daftar Pustaka.....	92
Lampiran	98

DAFTAR TABEL

Tabel 2.1 Literatur Review	9
Tabel 3.1 Tahapan Preprocessing Data	28
Tabel 3.2 Rangkaian Eksperimen pada Subset Data	34
Tabel 3.3 Tahapan Evaluasi <i>Threshold</i> Tuning pada Model XGBoost.....	35
Tabel 3.4 Tahapan Penerapan SMOTE dalam Pelatihan Model	36
Tabel 3.5 Alat Penelitian	37
Tabel 3.6 Urutan Jalan Penelitian.....	38
Tabel 4.1 Tahapan Evaluasi Baseline Model	46
Tabel 4.2 Pembagian Data dan Distribusi Kelas pada Data Latih dan Uji.....	49
Tabel 4.3 Tahapan Inisialisasi dan Pelatihan Awal Model XGBoost	51
Tabel 4.4 Tahapan Evaluasi Awal Model	52
Tabel 4.5 Tahapan Hyperparameter Tuning pada Model XGBoost.....	54
Tabel 4.6 Tahapan Evaluasi Model pada Subset Data 500	61
Tabel 4.7 <i>Confusion matrix</i> Subset Data 500	63
Tabel 4.8 Tahapan Evaluasi Model pada Subset Data 1000	64
Tabel 4.9 <i>Confusion matrix</i> Subset Data 1000	66
Tabel 4.10 Tahapan Evaluasi Model pada Subset Data 1000	66
Tabel 4.11 <i>Confusion matrix</i> Subset Data 2000	68
Tabel 4.12 Rekapitulasi dan visualisasi evaluasi subset data.....	69
Tabel 4.13 Ringkasan Evaluasi Subset.....	71
Tabel 4.14 Tahapan <i>Threshold Tuning</i>	73
Tabel 4.15 Hasil Evaluasi untuk Berbagai <i>Threshold</i>	74
Tabel 4.16 Tahapan dan Evaluasi Model dengan SMOTE	76
Tabel 4.17 Perbandingan Model XGBoost (<i>default</i>) vs <i>Threshold Tuning</i>	80
Tabel 4.18 Tabel Perbandingan Model XGBoost (<i>default</i>) vs SMOTE	80
Tabel 4.19 Fitur 10 <i>permission</i> beserta Label keterangan.....	82

DAFTAR GAMBAR

Gambar 2.1 Arsitektur Algoritma XGBoost	18
Gambar 3.1 Alur Metode Penelitian.....	26
Gambar 4.1 Pemisahan Fitur dan Label	44
Gambar 4.2 Jumlah Distribusi Kelas	45
Gambar 4.3 Hasil Evaluasi <i>Baseline</i>	47
Gambar 4.4 Rumus Perhitungan Rasio <i>scale_pos_weight</i>	48
Gambar 4.5 Distribusi Kelas Setelah Split	50
Gambar 4.6 Struktur Data Hasil Split.....	50
Gambar 4.7 Pelatihan Model XGBoost.....	52
Gambar 4.8 <i>Confusion matrix</i> Model Awal	53
Gambar 4.9 Pencarian Metrik terbaik.....	56
Gambar 4.10 Kurva ROC	59
Gambar 4.11 <i>Confusion matrix</i> Model Terbaik XGBoost	60
Gambar 4.12 Perbandingan metrik evaluasi masing-masing subset data.....	71
Gambar 4.13 Perbandingan <i>F1-Score</i> dan <i>Recall</i> per Subset.....	72
Gambar 4.14 Grafik Kompleksitas vs Waktu Komputasi	72
Gambar 4.15 ROC Curve pada SMOTE	79
Gambar 4.16 Tabel dataset dengan 10 fitur terpilih sebagai masukan model.....	83
Gambar 4.17 Visualisasi <i>decision tree</i> dari model XGBoost berdasarkan 10 fitur <i>permissions</i>	87

DAFTAR LAMPIRAN

Lampiran 1 Koding Eksperimen Project Deteksi Malware dengan XGBoost.....	98
Lampiran 2 Jadwal Penelitian	149
Lampiran 3 Lembar Bimbingan Dosen.....	150
Lampiran 4 Hasil Cek Plagiarisme	151

PERPUSTAKAAN
UNIVERSITAS JENDRAL ACHMAD YANI
YOGYAKARTA

DAFTAR SINGKATAN

APK	Android Package Kit
AUC	Area Under the Curve
CSV	Comma-Separated Values
FN	<i>False Negative</i>
FP	<i>False Positive</i>
ML	Machine Learning
OS	Operating System
SMOTE	Synthetic Minority Over-sampling Technique
TN	True Negative
TP	True Positive
XGBoost	Extreme Gradient Boosting
ROC	Receiver Operating Characteristic
<i>F1-Score</i>	Harmonic Mean of <i>Precision</i> and <i>Recall</i>
SVM	Support Vector Machine
GridSearchCV	Grid Search with Cross Validation
BSSN	Badan Siber dan Sandi Negara
Colab	Google Colaboratory
Mendeley Data	Platform Penyedia Dataset Penelitian