Klasifikasi Malware Android Dengan Menggunakan Metode XGBoost Algoritma

Proposal Tugas Akhir

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana
Informatika Universitas Muhammadiyah Malang

Taufik Abdul Aziz
202010370311434

Bidang Minat:
Sistem Keamanan Jaringan

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2024

LEMBAR PERSETUJUAN

Klasifikasi Malware Android Dengan Menggunakan Metode XGBoost Algoritma

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Prodi Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Taufik Abdul Aziz

202010370311434

Menyetujui,

Dosen 1	Dosen 2
Zamah Sari S.T., M.T	Christian Sri Kusuma Aditya S.Kom., M.Kom
NIDN 0708087701	NIDN 0727029101

LEMBAR PENGESAHAN

Klasifikasi Malware Android Dengan Menggunakan Metode XGBoost Algoritma

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
InformatikaUniversitas Muhammadiyah Malang

Disusun Oleh :

TAUFIK ABDUL AZIZ
202010370311434

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji pada tanggal 22 Oktober 2024

Menyetujui,

Dosen Penguji 1	Dosen Penguji 2
Vinna Rahmayanti S S.Si., M.Si	**Ir. Yufis Azhar S.Kom., M.Kom.**
NIP. 180306071990PNS.	**NIP. 10814100544PNS.**

Mengetahui,
Ketua Jurusan Informatika

Ir. Galih Wasis Wicaksono S.kom. M.Cs.
NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : **Taufik Abdul Aziz**
NIM : **202010370311434**
Fakultas/Jurusan : **Teknik/Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan "**KLASIFIKASI MALWARE ANDROID DENGAN MENGGUNAKAN METODE XGBOOST ALGORITMA**" beserta seluruh isinya dalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik Sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini atau da klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,

Dosen Pembimbing

Zamah Sari, S.T., M.T.
NIP. 10814100555PNS.

Malang, 23 September 2024

Yang bertandatangan

Taufik Abdul Aziz
202010370311434

Abstrak

Android, sistem operasi yang dikembangkan oleh Google, mendominasi pasar global dengan pangsa sebesar 71,8% pada akhir tahun 2023. Meskipun keberhasilan ini didorong oleh sifat open-source dan berbagai aplikasi di Google Play Store, Android juga menjadi target utama bagi serangan malware, dengan 97% dari semua serangan malware pada tahun 2022 ditujukan pada perangkat Android. Malware juga kian waktu terus mengalami peningkatan yang menjadikannya semakin sulit untuk dideteksi. Maka dari itu diperlukan metode deteksi yang andal. Pada bidang IT saat ini, machine learning telah menunjukkan hasil yang cukup efisien dalam mendeteksi malware. Penulis mengusulkan metode Algoritma XGBoost sebagai pendekatan klasifikasi malware Android. Dalam penelitian ini, dilakukan penerapan teknik Feature Selection, yaitu Recursive Feature Elimination (RFE) dan Multicollinearity Removal (MR), untuk mengurangi dimensi data dan meningkatkan performa model. Pengujian dilakukan dengan membandingkan kinerja model XGBoost sebelum dan sesudah penerapan Feature Selection. Hasil evaluasi menggunakan classification report dan confusion matrix menunjukkan bahwa model XGBoost yang menerapkan Feature Selection berhasil mencapai Validation Accuracy sebesar 98%, Detection Accuracy sebesar 98%, Precision sebesar 98%, Recall sebesar 98%, dan F1-Score sebesar 98%. Model tanpa Feature Selection hanya mencapai nilai 97% pada metrik yang sama.

Kata kunci: *Malware, Android, Machine Learning, Extreme Gradient Boosting, Klasifikasi*

Kata Pengantar

Puji syukur penulis ucapkan atas kehadirat Allah SWT yang atas Berkah dan Ridho-Nya penulis mampu menyelesaikan tugas akhir dengan judul "**Klasifikasi Malware Android Dengan Menggunakan Metode XGBoost Algoritma**", meskipun masih memiliki banyak kekurangan. Shalawat berangkai salam semoga tetap tercurah kepada junjungan kita Nabi Besar Muhammad SAW.

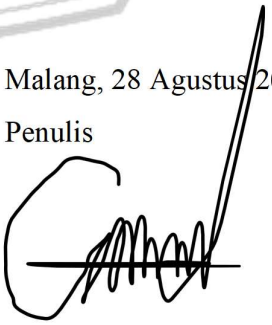
Penyusunan Tugas Akhir ini diajukan untuk memenuhi syarat akademis dalam rangka menyelesaikan Studi S1 Progam Studi Informatika di Fakultas Teknik Universitas Muhammadiyah Malang. Penyusunan Tugas Akhir ini tidak lepas dari bantuan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, dalam kesempatan ini ucapan syukur dan terima kasih penulis sampaikan kepada:

1. Allah SWT yang telah memberikan segala nikmat yang tak terhingga untuk penulis dan seluruh umat manusia. Serta, Nabi Muhammad SAW yang berkat perjuangannya membawa manusia dari zaman yang gelap menuju zaman yang terang benderang seperti sekarang.
2. Orang tua tersayang Ibu Yuningsih dan Bapak Munizar yang selalu memberi semangat, doa, nasehat, motivasi, serta materi yang tak akan pernah bisa penulis balas.
3. Bapak Zamah Sari, S.T., MT., selaku Dosen Pembimbing 1 dan Bapak Christian Sri Kusuma Aditya, S.Kom., M.Kom., selaku Dosen Pembimbing 2 yang selalu bersedia meluangkan waktu dan pikiran untuk memberikan bimbingan, arahan, serta saran dengan sabar untuk keberhasilan dan kebaikan Tugas Akhir ini.
4. Kepada keluarga dan teman-teman yang telah memberikan semangat dan masukan agar terselesaikannya rangkaian skripsi ini.

Penulis menyadari masih banyak kekurangan dalam penulisan Tugas Akhir ini. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun agar tulisan ini dapar berguna untuk perkembangan ilmu pengetahuan kedepannya.

Malang, 28 Agustus 2024

Penulis



Taufik Abdul Aziz

Daftar Isi

LEMBAR PERSETUJUAN ii
LEMBAR PENGESAHAN iii
LEMBAR PERNYATAAN iv
Abstrak v
Kata Pengantar vi
Daftar Isi vii
Daftar Gambar viii
Daftar Tabel ix
BAB 1 PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 3
1.3. Tujuan Penelitian 3
1.4. Batasan Masalah 3
BAB 2 LANDASAN TEORI 4
2.1. Penelitian Rujukan 4
2.2 . Android 8
2.3 Malware 9
2.4 Machine Learning 10
2.5 Klasifikasi 11
2.6 XGBoost 12
BAB 3 METHODOLOGI PENELITIAN 14
3.1 Alur Penelitian 14
3.2 Dataset 14
3.3 Preprocessing 17
3.4 Spliting Data 19
3.7 XGBoost Model 19
3.8 Result Evaluation 20
BAB 4 HASIL DAN PEMBAHASAN 22
4.1. Import Library 22
4.2 Visualisasi Data 22
4.3 Feature Selection 23
4.3.1 Recursive Feature Elimination 23
4.3.2 Multicollinearity Removal 24
4.3.3 Heatmap Feature Selection Comparison 24
4.4. Split Data 26
4.5. Model Evaluation 27
4.5.1. pengujian 27
BAB 5 KESIMPULAN 31
5.1 Kesimpulan 31
5.2 Saran 31
DAFTAR PUSTAKA 32

Daftar Gambar

Gambar 2. 1 Arsitektur XGBoost 13

Gambar 3. 1 Alur Penelitian 14

Gambar 4. 1 Library Yang Digunakan 22
Gambar 4. 2 Grafik Data Aplikasi Benign Dan Malware 23
Gambar 4. 3 Correlation Matrix sebelum Feature Selections 25
Gambar 4. 4 Correlation Matrix setelah Feature Selections 25
Gambar 4. 5 *Classification Report* 28
Gambar 4. 6 Confussion Matrix 28
Gambar 4. 7 Hasil klasifikasi Penelitian Mathur dkk. 29

Daftar Tabel

Tabel 2. 1 Penelitian Terdahulu. 4

Tabel 3. 1 Dataset Value. 16
Tabel 3. 2 Jenis-jenis Permissions 16
Tabel 3. 3 Evaluasi dari confusion matrix. 20
Tabel 3. 4 Rumus Penilaian 20

Tabel 4. 1 Recursive Feature Elimination 24
Tabel 4. 2 Multicollinearity Removal 24
Tabel 4. 3 Split Data 26
Tabel 4. 4 laporaan Classification Report 27

DAFTAR PUSTAKA

[1] "20 Android Statistics For 2024 (Market Share & Users)." Accessed: Jul. 19, 2024. [Online]. Available: https://www.demandsage.com/android-statistics/

[2] "Android Statistics 2024 - By Market Share, Useres and Revenue." Accessed: Jul. 19, 2024. [Online]. Available: https://www.enterpriseappstoday.com/stats/android-statistics.html#google_vignette

[3] "Attacks on mobile devices significantly increase in 2023 | Kaspersky." Accessed: Jul. 19, 2024. [Online]. Available: https://www.kaspersky.com/about/press-releases/2024_attacks-on-mobile-devices-significantly-increase-in-2023

[4] "Statistik malware Android dan perangkat lunak yang tidak diinginkan untuk Q1 2024 | Securelist." Accessed: Aug. 04, 2024. [Online]. Available: https://securelist.com/it-threat-evolution-q1-2024-mobile-statistics/112750/

[5] Y. Wanli Sitorus, P. Sukarno, S. Mandala, F. Informatika, and U. Telkom, "Analisis Deteksi Malware Android menggunakan metode Support Vector Machine & Random Forest," *e-Proceeding Eng.*, vol. 8, no. 6, pp. 12500–12518, 2021.

[6] R. B. Hadiprakoso, W. R. Aditya, and F. N. Pramitha, "Analisis Statis Deteksi Malware Android Menggunakan Algoritma Supervised Machine Learning," *Cyber Secur. dan Forensik Digit.*, vol. 5, no. 1, pp. 1–5, 2022, doi: 10.14421/csecurity.2022.5.1.3116.

[7] F. M. Darus, N. A. Ahmad, and A. F. M. Ariffin, "Android malware classification using XGBoost on data image pattern," *Proc. - 2019 IEEE Int. Conf. Internet Things Intell. Syst. IoTaIS 2019*, pp. 118–122, 2019, doi: 10.1109/IoTaIS47347.2019.8980412.

[8] J. Wang, B. Li, and Y. Zeng, "XGBoost-Based Android Malware Detection," pp. 268–272, 2017, doi: 10.1109/CIS.2017.00065.

[9] M. E. N. Et. al., "Malware Classification Using Xgboost With Vote Based Backward Feature Elimination Technique," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 10, pp. 5915–5923, 2021, doi: 10.17762/turcomat.v12i10.5412.

[10] J. Palša *et al.*, "MLMD—A Malware-Detecting Antivirus Tool Based on the XGBoost Machine Learning Algorithm," *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136672.

[11] A. Mathur, L. M. Podila, K. Kulkarni, Q. Niyaz, and A. Y. Javaid, "NATICUSdroid: A malware detection framework for Android using native and custom permissions," *J. Inf. Secur. Appl.*, vol. 58, p. 102696, May 2021, doi: 10.1016/J.JISA.2020.102696.

[12] N. Al Sarah, F. Y. Rifat, M. S. Hossain, and H. S. Narman, "An Efficient

Android Malware Prediction Using Ensemble machine learning algorithms," *Procedia Comput. Sci.*, vol. 191, no. 2019, pp. 184–191, 2021, doi: 10.1016/j.procs.2021.07.023.

[13] A. Meena, R. Karishma, B. Gayathri, and K. B. Hemapriya, "Android Malware Detection Using Extreme Gradient Boosting Algorithm," vol. 9, no. 4, 2023.

[14] L. Suhuan and H. Xiaojun, "Android malware detection based on logistic regression and XGBoost," *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, vol. 2019-Octob, pp. 528–532, 2019, doi: 10.1109/ICSESS47205.2019.9040851.

[15] "Android Architecture: Application Layers, Framework, Component." Accessed: Aug. 31, 2024. [Online]. Available: https://www.guru99.com/android-architecture.html

[16] "Android Architecture: A Comprehensive Overview Its Layers and Functions – Techporfit." Accessed: Aug. 31, 2024. [Online]. Available: https://techporfit.com/android-architecture/

[17] O. Aslan and R. Samet, "A Comprehensive Review on Malware Detection Approaches," *IEEE Access*, vol. 8, pp. 6249–6271, 2020, doi: 10.1109/ACCESS.2019.2963724.

[18] K. Liu, S. Xu, G. Xu, M. Zhang, D. Sun, and H. Liu, "A Review of Android Malware Detection Approaches Based on Machine Learning," *IEEE Access*, vol. 8, pp. 124579–124607, 2020, doi: 10.1109/ACCESS.2020.3006143.

[19] E. J. Sudarman and S. Budi, "Pengembangan Model Kecerdasan Mesin Extreme Gradient Boosting untuk Prediksi Keberhasilan Studi Mahasiswa," *J. Strateg.*, vol. 5, no. 2, pp. 297–314, 2023.

[20] "Manifest.permission | Android Developers." Accessed: Jul. 21, 2024. [Online]. Available: https://developer.android.com/reference/android/Manifest.permission

[21] "Data Wrangling for Machine Learning | StreamSets." Accessed: Jul. 21, 2024. [Online]. Available: https://www.softwareag.com/en_corporate/blog/streamsets/data-wrangling-for-machine-learning.html

[22] J. A. Ramírez-Hernández and E. Fernandez, "Enhanced recursive feature elimination," *Proc. - 6th Int. Conf. Mach. Learn. Appl. ICMLA 2007*, pp. 330–335, 2007, doi: 10.1109/ICMLA.2007.35.

[23] A. N. Iman, M. T. Avon Budiyono, S.T., and M. T. Ahmad Almaarif, S.Kom., "Analisis Malware Pada Sistem Operasi Android Menggunakan Permission-Based Malware Analysis in Android Operation System Using Permission-Based," *Angew. Chemie Int. Ed. 6(11), 951–952.*, vol. 6, no. Mi, pp. 5–24, 1967.

UNIVERSITAS
MUHAMMADIYAH
MALANG

FAKULTAS TEKNIK

INFORMATIKA
informatika.umm.ac.id | informatika@umm.ac.id

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Taufik Abdul Aziz

NIM : 202010370311434

Judul TA : Klasifikasi Malware Android dengan Menggunakan Metode XGBoost Algoritma

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	6%
2.	Bab 2 – Daftar Pustaka	25 %	25%
3.	Bab 3 – Analisis dan Perancangan	25 %	3%
4.	Bab 4 – Implementasi dan Pengujian	15 %	0%
5.	Bab 5 – Kesimpulan dan Saran	5 %	0%
6.	Makalah Tugas Akhir	20%	20%

) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

***) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)**

Mengetahui,

Pemeriksa (Staff TU)

(.....)

Kampus I
Jl. Bandung 1 Malang, Jawa Timur
P: +62 341 551 253 (Hunting)
F: +62 341 460 435

Kampus II
Jl. Bendungan Sutami No.188 Malang, Jawa Timur
P: +62 341 551 149 (Hunting)
F: +62 341 582 060

Kampus III
Jl. Raya Tlogomas No.246 Malang, Jawa Timur
P: +62 341 464 318 (Hunting)
F: +62 341 460 435
E: webmaster@umm.ac.id