Klasifikasi Malware Android dengan Menggunakan Metode CatBoost Algoritma

LAPORAN TUGAS AKHIR

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

Yusuf Irsyaduddin
(201910370311348)

Bidang Minat
(Sistem Keamanan Jaringan)

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2023

LEMBAR PERSETUJUAN

Klasifikasi Malware Android Dengan Menggunakan Metode CatBoost Algoritma

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Menyetujui,
Malang, *30 Desember 2023*

Dosen Pembimbing 1	Dosen Pembimbing 2
Ir Denar Regata Akbi S.Kom., M.Kom.	**Didih Rizki Chandranegara S.kom., M.Kom**
NIP. 10816120591PNS.	**NIP. 180302101992PNS.**

LEMBAR PENGESAHAN

Klasifikasi Malware Android Dengan Menggunakan Metode CatBoost Algoritma

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
InformatikaUniversitas Muhammadiyah Malang

Disusun Oleh :

YUSUF IRSYADUDDIN

201910370311348

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji pada tanggal 30 Desember 2023

Menyetujui,

Dosen Penguji 1

Christian Sri Kusuma Aditya S.Kom., M.Kom

NIP. 180327021991PNS.

Dosen Penguji 2

Luqman Hakim S.Kom., M.Kom.

NIP. 10819030658PNS.

Mengetahui,
Ketua Jurusan Informatika

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

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : **Yusuf Irsyaduddin**

NIM : **201910370311348**

FAK./JUR. : **Informatika**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **"Klasifikasi Malware Android Dengan Menggunakan Metode CatBoost Algoritma"** beserta seluruh isinya adalah 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 ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing

Ir Denar Regata Akbi S.Kom., M.Kom.

Malang, 30 Desember 2023
Yang Membuat Pernyataan

Yusuf Irsyaduddin

ABSTRAK

Pada tahun 2008, Android diperkenalkan sebagai proyek sumber terbuka yang populer karena kemampuan penyesuaian dan persyaratan perangkat keras yang rendah. Statistik pertengahan tahun 2021 dari *GlobalStat Counter* menunjukkan bahwa android mendominasi pasar sistem operasi seluler dengan 72,74%. Meskipun popularitas, android menjadi target serangan malware dalam konteks kejahatan cyber. Permasalahan tersebut yang mendorong penelitian ini dilakukan dengan tujuan kebutuhan untuk mengidentifikasi dan mengklasifikasikan malware android yang terus menerus semakin berkembang dengan menerapkan logika machine learning, khususnya menggunakan metode *CatBoost*. Metode ini dipilih berdasarkan keefektifannya dalam penelitian sebelumnya yang telah terbukti memberikan akurasi tinggi. Evaluasi kinerja melibatkan perbandingan antara *CatBoost* dan beberapa metode peneliti sebelumnya, termasuk *KNN (K-Nearest Neighbors), SVM (Support Vector Machine), LR (Logistic Regression), RF (Random Forest), ET (Extra Trees), XG (XGBoost), AB (Adaboost),* dan *BG (Bagging),* menggunakan metrik umum seperti *Validation Accuracy*, *Detection Accuracy,* dan *F1-Score*. Hasil penelitian menunjukkan bahwa *CatBoost* berhasil mencapai *Validation Accuracy* sebesar 96,66%, *Detection Accuracy* 96,87%, dan *F1-Score* sebesar 96,81% menempatkannya di posisi yang bersaing dengan sebagian besar metode lain, kecuali *RF (Random Forest)*. Keunggulan konsisten CatBoost dalam perbandingan ini menunjukkan potensi sebagai solusi efektif dan konsisten dalam deteksi dan klasifikasi malware Android.

Kata Kunci : ***Android, Malware, Machine learning, CatBoost.***

ABSTRACT

In 2008, Android was introduced as a popular open source project due to its customizability and low hardware requirements. Mid-2021 statistics from *GlobalStat Counter* shows that Android dominates the mobile operating system market with 72.74%. Despite its popularity, Android is becoming a target for malware attacks in the context of cyber crime. This problem prompted this research to be carried out with the aim of identifying and classifying Android malware which is continuously developing by applying machine learning logic, especially using the method*CatBoost*. This method was chosen based on its effectiveness in previous research which has been proven to provide high accuracy. Performance evaluation involves comparisons between*CatBoost* and several previous researchers' methods, incl*KNN (K-Nearest Neighbors), SVM (Support Vector Machine), LR (Logistic Regression), RF (Random Forest), ET (Extra Trees), XG (XGBoost), AB (Adaboost),* and *BG (Bagging)*, using common metrics such as*Validation Accuracy*, *Detection Accuracy*, and *F1-Score*. The research results show that*CatBoost* managed to achieve*Validation Accuracy* amounting to 96.66%,*Detection Accuracy* 96,87%, and*F1-Score* of 96.81% puts it in a competitive position with most other methods, except*RF (Random Forest)*. CatBoost's consistent superiority in this comparison shows its potential as an effective and consistent solution in Android malware detection and classification.

Keywords :*Android, Malware, Machine learning, CatBoost.*

LEMBAR PERSEMBAHAN

Puji syukur saya panjatkan kepada Allah SWT, karena atas berkat, rahmat dan hidayahNya, penulis dapat menyelesaikan skripsi ini dengan judul "Klasifikasi Malware Android dengan Menggunakan Metode CatBoost Algoritma". Dalam proses penyusunan skripsi ini tidak terlepas dari dukungan dan bantuan dari semua pihak, sehingga pada kesempatan ini, penulis tidak lupa mengucapkan banyak terima kasih kepada semua pihak yang banyak membantu dalam proses penyusunan skripsi ini. Khususnya ucapan terima kasih kepada:

1. Allah SWT atas semua Keridhoan-Nya dan Izin-Nya sehingga penulis mampu menyelesaikan tugas akhir.
2. Ir Denar Regata Akbi S.Kom., M.Kom selaku dosen pembimbing tugas akhir pertama dan Bapak Didih Rizki Chandranegara, S.Kom., M.Kom selaku dosen pembimbing kedua.
3. Bapak Dekan Fakultas Teknik Universitas Muhammadiyah Malang.
4. Bapak Ketua Jurusan Teknik Informatika Universitas Muhammadiyah Malang.
5. Jajaran Dosen Program Studi Informatika Universitas Muhammadiyah Malang.
6. Kepada kedua orang tua yang telah mendukung saya hingga sejauh ini.
7. Abdur, Hadid, Jalal, Moris, dan Rafi yang telah menjadi sahabat sekaligus teman belajar selama perkuliahan sampai skripsi, serta selalu menopang satu sama lain.
8. Seluruh rekan Mahasiswa Informatika angkatan 2019, khususnya kelas G atas kebersamaannya selama kuliah.

Malang, 4 Desember 2023

YUSUF IRSYADUDDIN

KATA PENGANTAR

Alhamdulillahi Robbil'alamin dengan memanjatkan puji syukur atas kehadirat Allah SWT, serta shalawat serta salam kepada junjungan kita Nabi Muhammad SAW, sehingga dengan ridho dan rahmat-Nya peneliti dapat menyelesaikan tugas akhir yang berjudul: "KLASIFIKASI MALWARE ANDROID DENGAN MENGGUNAKAN METODE CATBOOST ALGORITMA".

Di dalam penulisan ini disajikan pokok-pokok bahasan yang meliputi model yang diusulkan terhadap klasifikasi berbasis machine learning untuk malware android. Tugas Akhir ini ditulis untuk memenuhi persyaratan agar menerima gelar sarjana di bidang Informatika Fakultas Teknik Universitas Muhammadiyah Malang.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 4 Desember 2023

YUSUF IRSYADUDDIN

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR PUSTAKA

[1] M. M. Alani and A. I. Awad, "PAIRED: An Explainable Lightweight Android Malware Detection System," IEEE Access, vol. 10, no. July, pp. 73214–73228, 2022, doi: 10.1109/ACCESS.2022.3189645.

[2] [Net Marketshare, "Operating System Market Share," [Online]. Available: https://netmarketshare.com/operating-system-market-share.aspx. [Access on 23 June 2023].

[3] Universitas Terbuka. (2021). Keamanan dan Ancaman pada Cyberspace (1st ed.). https://pustaka.ut.ac.id/lib/wp-content/uploads/pdfmk/MSIM4404-M1.pdf

[4] Palumbo, L. Sayfullina, D. Komashinskiy, E. Eirola, and J. Karhumen, "A Pragmatic Android Malware Detection Procedure," Computers & Security, July 2017

[5] Urcuqui-Lopez, C., & Navarro Cadavid, A. (2016). Framework for malware analysis in Android. Sistemas y Telemática, 14(37), 45–56. https://doi.org/10.18046/syt.v14i37.2241

[6] 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, no. January, p. 102696, 2021, doi: 10.1016/j.jisa.2020.102696.

[7] Jiong, Wang & Li, Boquan & Zeng, Yuwei. (2017). XGBoost-Based Android Malware Detection. 268-272. 10.1109/CIS.2017.00065.

[8] C. Prabhavathi, A. Mahesh, S. V. Ajay, and R. Mythili, "Malware Prediction Using XGBOOST and CATBOOST," J. Eng. Sci., vol. 13, no. 05, pp. 620–626, 2022, [Online]. Available: https://jespublication.com/upload/2022-V13I5085.pdf

[9] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *J. Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00369-8.

[10] Yandex. (2023). Catboost, catboost enables fast gradient boosting on

decision trees using gpus. Diakses pada 9 Januari 2023 dari https://catboost.ai/news/catboost- enables-fast-gradient-boosting-on-decision-trees-using-gpus.

[11] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," pp. 1–7, 2018, [Online]. Available: http://arxiv.org/abs/1810.11363

[12] Yandex. (2023). Catboost, Parameter tuning. Diakses pada 6 Oktober 2023 dari https://catboost.ai/en/docs/concepts/parameter-tuning.

[13] P. Agrawal and B. Trivedi, "Evaluating Machine Learning Classifiers to detect Android Malware," *2020 IEEE Int. Conf. Innov. Technol. INOCON 2020*, pp. 1–6, 2020, doi: 10.1109/INOCON50539.2020.9298290.

[14] S. Ben Jabeur, C. Gharib, S. Mefteh-Wali, and W. Ben Arfi, "CatBoost model and artificial intelligence techniques for corporate failure prediction," *Technol. Forecast. Soc. Change*, vol. 166, no. October 2020, p. 120658, 2021, doi: 10.1016/j.techfore.2021.120658.

[15] N. Setiawan, "Kasus Kejahatan Siber Pada Telepon Seluler Android," Cyber Secur. dan Forensik Digit., vol. 2, no. 1, pp. 24–29, 2019, doi: 10.14421/csecurity.2019.2.1.1420.

[16] Febrianto, A. F., Budiyono, A., & Almaarif, A. (2019). Analisis Malware Pada Sistem Operasi Android Menggunakan Metode Network Traffic Analysis. eProceedings of Engineering.

[17] Iman, A. N., Budiyono, A., & Almaarif, A. (2019). Analisis Malware Pada Sistem Operasi Android Menggunakan Permission-based. eProceedings of Engineering.

[18] S. Sinambela, A. R. Pangestu, and R. Feriyanto, "Analisis Aplikasi Malware pada Android dengan Metode Statik," J. Ilm. Ilk. - Ilmu Komput. Inform., vol. 3, no. 2, pp. 88–94, 2020, doi: 10.47324/ilkominfo.v3i2.101.

[19] IBM. "What is machine learning?". IBM. https://www.ibm.com/topics/machine- learning. Diakses pada 2 Juli 2023.

[20] Katrina, W. “A guide to the types of machine learning alorithms and their applications”. SAS. https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html. Diakses pada 2 Juli 2023.

[21] P. R. Sihombing and I. F. Yuliati, “Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 417–426, 2021, doi: 10.30812/matrik.v20i2.1174.

[22] Adebayo, Samuel.” How CATBOOST Algorithm Works in Machine Learnig”. *Dataaspirant*, https://dataaspirant.com/catboost-algorithm/#t-1609567161983. Diakses pada 3 Juli 2023.

[23] Esri. “How CatBoost algorithm works”. Esri. https://pro.arcgis.com/en/pro-app/latest/tool-reference/geoai/how-catboost-works.htm. Diakses pada 3 Juli 2023.

[24] Android Developers. “Manifest.permission”. Developer Android.https://developer.android.com/reference/android/Manifest.permission . Diakses pada 26 Agustus 2023.

[25] Sean, A. “Data Wrangling for Machine Learning”. StreamSets. https://streamsets.com/blog/data-wrangling-for-machine-learning/. Diakses pada 6 Juli 2023.

[26] J. Li *et al.*, “Feature selection: A data perspective,” *ACM Comput. Surv.*, vol. 50, no. 6, 2017, doi: 10.1145/3136625.

UNIVERSITAS
MUHAMMADIYAH
MALANG

FAKULTAS TEKNIK

INFORMATIKA

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

FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : **YUSUF IRSYADUDDIN**

NIM : **201910370311348**

Judul TA : **Klasifikasi Malware Android dengan Menggunakan Metode CatBoost Algoritma**

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	5 %
2.	Bab 2 – Daftar Pustaka	25 %	2 %
3.	Bab 3 – Analisis dan Perancangan	25 %	2 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	2 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	0 %
6.	Makalah Tugas Akhir	20%	7 %

**) 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