KLASIFIKASI ADWARE MALWARE PADA ANDROID DENGAN METODE *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Strata 1

OLEH:

NOVIT HARDIANTO
09011281520086

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN

KLASIFIKASI ADWARE MALWARE PADA ANDROID DENGAN METODE *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

OLEH :

Novit Hardianto
09011281520086

Indralaya, 31 Desember 2020

Pembimbing I	Pembimbing II
Deris Stiawan, M.T., Ph.D.	**Ahmad Heryanto,S.Kom., M.T.**
NIP. 197806172006041002	**NIP.198701222015041002**

Mengetahui,
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T
NIP. 196612032006041001

HALAMAN PERSETUJUAN

Pada hari Kamis 31 Desember 2020 telah dilaksanakan ujian sidang tugas akhir oleh Sarjana Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama	: Novit Hardianto
NIM	09011281520086
Judul	: **KLASIFIKASI ADWARE MALWARE PADA ANDROID DENGAN METODE RANDOM FOREST**

Tim Penguji :

1. **Penguji I**

   Ahmad Zarkasi, M.T . (.....)

Mengetahui,
Ketua Jurusan Sistem Komputer

Dr.Ir.H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Novit Hardianto

NIM : 09011281520086

Judul : **Klasifikasi Adware Malware Pada Android Dengan Metode *Random Forest***

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari universitas Sriwijaya. Demikian, pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan.

Indralaya, 31 Desember 2020

(Novit Hardianto)

HALAMAN PERSEMBAHAN

"Tuhanmu tiada meninggalkanmu dan tiada (pula) benci kepadamu."
(Q.S. Ad-Duha 93:3)

"pantang pulang sebelum S.kom"
-Bapak

"jangan mudah menyerah coba saja dulu. Siapa tahu rezekinya datang dari situ."
-Mamak

"ada orang- orang yang berdoa siang malam untuk kamu jangan patahkan itu, Jangan menyalahkan kondisi dan keadaan dengan buruknya manejemen kamu , dan semoga kamu jadi lebih baik kedepan"
-Pak deris setiawan

"tingalkan sejenak kesenanganmu demi kesuksesanmu kelak"
-novit hardianto

Kupersembahkan khusus untuk seorang ibu bernama hartini dan seorang ayah bernama sardi . Mak, pak, terimakasih telah memperjuangkanku sejauh ini. Rasa bahagia dan banggamu adalah hal yang selalu kuharapkan.
Dan untuk kakak perempuanku : wiwit defratini ningrum, terimakasih karena kamu telah mengajarkanku arti kerja keras, kemandirian dan sikap pantang menyerah dalam hidup.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan proposal tugas akhir dengan judul "**Klasifikasi Adware Malware Pada Android Dengan Metode *Random Forest***". Pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah SWT dan terima kasih kepada yang terhormat :

1. Kedua Orang Tua, Bapak Sardi, Ibu Hartini, Kakak perempuan, Wiwit Defratini Ningrum, S.P, dan pacar saya Dina prihatilia yang telah memberikan do'a dan dukungannya serta memberikan motivasi untuk tetap selalu berusaha dan Tawakal.
2. Bapak Jaidan Jauhari, S.Pd., M.T., Selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Dr. Ir. H. Sukemi, M. T.. Selaku Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
4. Bapak Deris Stiawan, M.T., Ph.D.,Selaku PembimbingI Tugas Akhir di Jurusan Sistem Komputer.
5. Bapak Ahmad Heryanto, S.Kom., M.T.,Selaku Pembimbing II Tugas Akhir di Jurusan Sistem Komputer.
6. Seluruh Dosen Jurusan Sistem Komputer Fasilkom Universitas Sriwijaya.
7. Mbak Renny virgasari selaku admin jurusan Sistem Komputer.
8. Rekan-rekan seperjuangan Sistem Komputer, dyah citra soraya Aprilia rahmayanti, fernanda, tiara juwita dan angkatan 2015 gank kapak. S.kom

Penulis menyadari bahwa proposal ini masih jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun sangat penulis harapkan sebagai bahan acuan dan perbaikan untuk penulis dalam menyempurnakan laporan ini.

Indralaya, 31 Desember 2020

Novit hardianto

ADWARE MALWARE CLASSIFICATION ON ANDROID USING RANDOM FOREST METHOD

Novit hardianto (09011281520086)
Department of Computer Systems, Faculty of Computer Science,
Sriwijaya University
E-mail: novithardianto86@gmail.com

Abstrak

The development of technology triggers the development of malicious files called malware. Malware is software that is explicitly designed with the aim of finding weaknesses or even damaging software or operating systems. In this study, the dowgin and benign malware classification was carried out using the Random Forest algorithm method by comparing weka data and spyder programs. The dataset used in this study is the CICAndMal2017 csv (Comma Separated Values) category with the dowgin type in this dataset has 1197 for 53% dowgin data and 792 begign data or 47% where this dataset has 85 attributes. After the classification, the accuracy value for the accuracy value is 0.998% and the OOB Error value is 0.16%, while using the Random Forest method the accuracy value for the spyder program is 0.891% and the OOB Error value is 0.108%.

keywords: adware malware, weka, spyder, Random Forest

KLASIFIKASI ADWARE MALWARE PADA ANDROID DENGAN METODE *RANDOM FOREST*

Novit hardianto (09011281520086)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer Universitas Sriwijaya

E-mail : novithardianto86@gmail.com

Abstrak

Semakin berkembangnya teknologi memicu berkembangnya file-file jahat yang disebut malware. Malware merupakan perangkat lunak yang secara eksplisit didesain dengan tujuan mencari kelemahan atau bahkan merusak *software* atau sistem operasi. Pada penelitian ini dilakukan klasifikasi malware dowgin dan benign menggunakan metode algoritma Random Forest dengan cara membandingkan data weka dan program *spyder*. Dataset yang digunakan pada penelitian ini yaitu *CICAndMa*l2017 kategori csv (*Comma Separated Values*) dengan jenis dowgin pada dataset ini memiliki 1197 untuk data dowgin 53% dan 792 data begign atau 47% dimana dataset ini memliki 85 atribut. Setelah dilakukan klasifikasi didapatkan hasil nilai akurasinnya untuk weka nilai *accuracy* yaitu 0,998% dan nilai OOB Errornya 0,16% sedangkan dengan mengunakan metode Random Forest untuk nilai akurasi pada program *spyder* yaitu 0,891% dan nilai OOB Errornya 0.108%.

Kata Kunci : *Adware Malware , Weka , Spyder, Random Forest*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1. Latar Belakang

Pada penelitian ini menjelaskan bahwa android merupakan salah satu sistem operasi yang banyak digunakan pada saat ini. Smartphone berbasis Android sangat membantu user dalam melakukan berbagai aktivitas seperti berbelanja online, berkomunikasi, presentasi, dan masih banyak lagi. Android adalah salah satu sistem operasi berbasis linux. Kelebihan Android dibanding sistem operasi mobile phone atau smartphone lainnya adalah Android bersifat open source code sehingga memudahkan para pengembang untuk menciptakan dan memodifikasi aplikasi atau fitur – fitur yang belum ada pada sistem operasi Android sesuai dengan keinginan mereka sendiri. Semakin berkembangnya teknologi pada masa sekarang juga memicu berkembangkan file-file jahat yang biasa disebut sebagai malware.

Penelitian selanjutnya malware merupakan program komputer yang diciptakan dengan tujuan mencari kelemahan atau bahkan merusak software atau sistem operasi, malware biasanya menyusup kedalam suatu program atau aplikasi dan merusak sistem Android bahkan ada yang bisa mencuri file-file penting yang ada dalam perangkat smartphone, sehingga banyak pengguna yang merasa dirugikan. Malware merupakan perangkat lunak yang secara eksplisit didesain untuk melakukan aktifitas berbahaya atau perusak perangkat lunak lainnya, seperti Trojan, Virus, Spyware, adware, dan Exploit.

Pada penelitian yang dilakukan menggunakan lebih dari 270.000 sampel malware dan 837 sampel adware yang dijalankan di dalam program spyder kemudian diklasifikasikan menggunakan Random Forests dan representasi fitur baru untuk deteksi dan klasifikasi keluarga malware, mampu mencapai deteksi malware dengan rata-rata TPR dan PPV masing-masing 0,981, sedangkan klasifikasi keluarga malware mencapai rata-rata TPR,PPV,AUC masing-masing 0,860, 0,867 dan 0,977. Uji coba tersebut menunjukan deteksi malware menggunakan teknik analisis dinamik dengan sangat efisien untuk deteksi dan mengklasifikasi malware.

Teorema bayes merupakan *teorema* yang digunakan dalam statistic untuk menghitung peluang dari sebuah hipotesis, dengan cara menghitung peluang suatu kelas berdasarkan attribute yang akan digunakan dan menentukan kelas probabilitas untuk melihat probabilitas yang tertingi. Dalam penelitian menyebutkan *Naïve bayes* merupakan metode yang melakukan klasifikasi data berdasarkan attribute data yang dinyatakan dengan $x=(x1,x2,x3,..,xn)$.

Penggunaan *Naïve Bayes* ini hanya membutuhkan jumlah data training yang kecil untuk menentukan parameter diperlukan dalam klasifikasi nya[11].

Dari beberapa pembahasan diatas, penulis bermaksud melakukan penelitian klasifikasi terhadap malware adware dengan mengunakan metode naive bayes untuk dapat mengenali pola serangan malware adware dengan data normal

1.2. Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Melakukan klasifikasi malware dowgin dan benign menggunakan algoritma Random Forest.
2. Menerapkan metode Random forest sebagai feature selection pada malware Adware
3. Mengklasifikasi malware berdasarkan prilaku dengan mengunakan algoritma Random forest classifiers

1.3. Manfaat

1. Dapat mengklasifikasikan data yang merupakan malware dowgin dan benign.
2. Dapat menerapkan algoritma untuk klasifikasi data dan feature selection pada malware dan non-malware.
3. Mengetahui tingkat akurasi algoritma *Random forest classifiers* dalam mengklasifikasi *Malware*.

1.4. Rumusan dan Batasan Masalah

Rumusan masalah yang akan diambil dari penelitian ini adalah bagaimana hasil klasifikasi dari penerapan algoritma Random Forest sehingga ditemukan akurasi yang lebih baik dari penelitian sebelumnya

1.Bagaimana membangun sistem analisis dinamis dengan mnggunakan program spyder untuk menganalisa prilaku *malware* secara otomatis.

2. Bagaimana mengklasifikasi Malware dan Normal file dengan algoritma *Random forest classifiers*

1.5. Batasan masalah

Agar penelitian mengarah pada pemaparan yang diharapkan, maka diperlukan batasan masalah dalam penelitian. Adapun batasan masalah dalam penelitian ini, adalah :

1. Dataset yang digunakan dalam penelitian ini berbatas pada dataset yang berasal dari Canadian *Institute for Cybersecurity* (CIC) adalah CICAndMal2017 pada kategori *adware* yaitu *dowgin* dan *Benign.*

2. Mengklasifikasi Malware dan Normal File dengan program spyder dan algoritma *Random forest classifiers.*

3. Tidak membahas bagaiman malware dapat masuk kedalam sistem komputer.

1.5. Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini akan melewati beberapa tahapan sebagai berikut:

1. Tahap Pertama (Studi Pustaka/Literatur)

   Tahap ini dilakukan setelah mslah yang akan dibahas telah sesuai dan relevan untuk diangkat sebagai penelitian, dengan mencari referensi atau literature pada Keyword yang di angkat dari judul yang bertujuan untuk menunjang pada penelitian yang dilakukan.

2. Tahap Kedua (Perancangan Sistem)

   Tahap ini ialah tahap perancangan sistem yang dibuat berdasarkan perumusan masalah yang dicari dalam penelitian. Dalam tahap ini, membangun dan menerapkan metode yang akan digunakan serta menyiapkan hardwaredan software serta melakukan konfigurasi ataupun menulis code untuk penerapan pada tugas akhir.

3. Tahap Ketiga (Pengujian)

   Pada tahap ini, data yang diperoleh adalah dataset Canadian *Institute for Cybersecurity* (CIC) yaitu CICAndMal2017 di kategori adware , *Dowgin* dan Benign

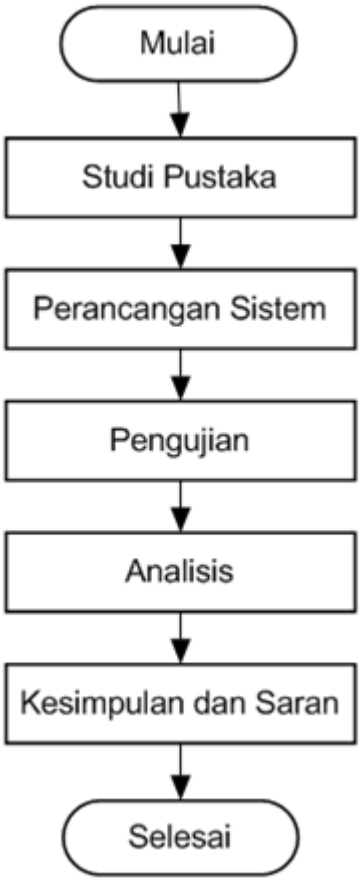
4. Tahap Keempat (Analisis)

   Tahap ini dilakukan dengan mengambil data dan menganalisa data yang didapatkan dari tahap ketiga yaitu tahap pengujian yang bertujuan apakah sistem berjalan dengan baik atau masih ada kekurangan, sehingga didapatkan data yang objektif.

5. Tahap Kelima (Kesimpulan dan Saran)

   Tahap ini akan dirumuskan suatu kesimpulan berdasarkan permaslahan, studi pustaka, metodologi penelitian dan analisa hasi pengujian, serta saran untuk penulis selanjutnya jika akan dijadikan bahan referensi

Tampilan gambaran proeses penelitian sebagai berikut :



1.6. Sistematika Penulisan

Penyusunan laporan tugas akhir ini, penulis membuat sistematika penulisan agar mempermudah mengetahui isi dari setiap bab yang dibuat pada laporan tugas akhir ini. Adapun seitematika penulisan laporan tugas akhir sebagai berikut :

BAB I. PENDAHULUAN

Bab ini akan menjelaskan tentang latar belakang masalah, tujuan danmanfaat, perumusan masalah, batasan masalah, metodologi penelitian,serta sistematika penulisan.

BAB II.TINJAUAN PUSTAKA

Bab ini berisi dasar teori dari penelitian terkait dengan malware adware, dengan mengunakan metode *Random Forest*, dan yang berkaitan langsung dengan penelitian

BAB III.METODELOGI

Bab ini akan menjelaskan tentang langkah-langkah (metodologi) perancangan sistem pada tugas akhir ini..

BAB IV.PENGUJIAN DAN ANALISA

Bab ini akan menjelaskan tentang hasil dari pengujian yang telah dilakukan, dari hasil tersebut akan dilakukan analisa agar mendapatkan data yang akurat.

BAB V. KESIMPULAN

Bab ini akan menjelaskan tentang kesimpulan yang didapat dari data penelitian yang telah dilakukan. Dan saran yang diharapkan dapat membuat penelitian ini dikembangkan lebih baik.

DAFTAR PUSTAKA

[1] A. H. Lashkari, A. F. A. Kadir, L. Taheri, and A. A. Ghorbani, “Toward Developing a Systematic Approach to Generate Benchmark Android Malware Datasets and Classification,” *2018 Int. Carnahan Conf. Secur. Technol.*, no. Cic, pp. 1–7, 2018.

[2] K. Allix, Q. Jerome, T. F. Bissyande, J. Klein, R. State, and Y. Le Traon, “A forensic analysis of android malware - How is malware written and how it could be detected?,” *Proc. - Int. Comput. Softw. Appl. Conf.*, pp. 384–393, 2014, doi: 10.1109/COMPSAC.2014.61.

[3] J. Milosevic, A. Ferrante, and M. Malek, “MalAware: Effective and Efficient Run-Time Mobile Malware Detector,” *Proc. - 2016 IEEE 14th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2016, 2016 IEEE 14th Int. Conf. Pervasive Intell. Comput. PICom 2016, 2016 IEEE 2nd Int. Conf. Big Data Intell. Comput. DataCom 2016 2016 IEEE Cyber Sci. Technol. Congr. CyberSciTech 2016, DASC-PICom-DataCom-CyberSciTech 2016*, pp. 270–277, 2016, doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2016.65.

[4] D. R. Septiani, N. Widiyasono, and H. Mubarok, “Investigasi Serangan Malware Njrat Pada PC,” *J. Edukasi dan Penelit. Inform.*, vol. 2, no. 2, pp. 123–128, 2016, doi: 10.26418/jp.v2i2.16736.

[5] M. Ramzan, “Comparing and evaluating the performance of WEKA classifiers on critical diseases,” *India Int. Conf. Inf. Process. IICIP 2016 - Proc.*, pp. 1–4, 2017, doi: 10.1109/IICIP.2016.7975309.

[6] N. K. Suchetha, A. Nikhil, and P. Hrudya, “Comparing the Wrapper Feature Selection Evaluators on Twitter Sentiment Classification,” *2019 Int. Conf. Comput. Intell. Data Sci.*, pp. 1–6, 2019.

[7] M. Peker, “A Novel Hybrid Method for Determining the Depth Combining ReliefF Feature Selection and Random Forest Algorithm (ReliefF + RF),” 2015.

[8] F. A. Kurniawan, A. P. Kurniati, and R. Tree, “ANALISIS DAN IMPLEMENTASI RANDOM FOREST DAN CLASSIFICATION DAN REGRESSION TREE (CART) UNTUK KLASIFIKASI PADA MISUSE INTRUSION DETECTION SYSTEM,” 2011.

[9] V. Mhetre and M. Nagar, “Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA,” *Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017*, vol. 2018-January, no. Iccmc, pp. 475–479, 2018, doi: 10.1109/ICCMC.2017.8282735.

[10] A. H. Muhammad, B. Sugiantoro, A. Luthfi, M. Teknik, I. Universitas, and I. Indonesia, “Metode Klasifikasi Dan Analisis Karakteristik Malware Menggunakan Konsep Ontologi,” no. 1, 2004.

[11] I. Charalampopoulos and I. Anagnostopoulos, “A comparable study employing weka clustering/classification algorithms for web page classification,” *Proc. - 2011 Panhellenic Conf. Informatics, PCI 2011*, pp. 235–239, 2011, doi: 10.1109/PCI.2011.52.

[12] E. S. Lamdompak Sistem Komputer and F. Ilmu Komputer, “Klasifikasi Malware Trojan Ransomware Dengan Algoritma Support Vector Machine (SVM),” vol. 2, no. 1, pp. 122–127, 2016, [Online]. Available: http://ars.ilkom.unsri.ac.id.

[13] R. Ferdiana, F. Jatmiko, D. D. Purwanti, A. S. T. Ayu, and W. F. Dicka, “Dataset Indonesia untuk Analisis Sentimen,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 8, no. 4, p. 334, 2019, doi: 10.22146/jnteti.v8i4.533.