KLASIFIKASI *ANDROID MALWARE* MENGGUNAKAN ALGORITMA *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Strata 1

OLEH:

DYAH CITRA SORAYA
09011281520107

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2020

LEMBAR PENGESAHAN

KLASIFIKASI *ANDROID MALWARE* MENGGUNAKAN ALGORITMA *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *RANDOM FOREST*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Strata 1

Oleh :

DYAH CITRA SORAYA
09011281520107

Indralaya, Januari 2021
Mengetahui,
Ketua Jurusan Sistem Komputer

Pembimbing

Deris Stiawan, M.T., PH.D.
NIP. 198106162012121003

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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:

Hari : Kamis

Tanggal : 31 Desember 2020

Tim Penguji :

1. Ketua : Ahmad Heryanto, M.T

2. Anggota I : Rahmad Fadli Isnanto, M.Sc

Mengetahui,

Ketua Jurusan Sistem Komputer

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

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : Dyah Citra Soraya

NIM : 09011281520107

Judul : Klasifikasi *Android Malware* Menggunakan Algoritma *Principal Component Analysis* (PCA) dan *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, Januari 2021

METERAI TEMPEL
C6900AHF794508185
6000
ENAM RIBU RUPIAH

Dyah Citra Soraya
NIM. 09011281520107

HALAMAN PERSEMBAHAN

"Allah tidak membebankan seseorang melainkan sesuai dengan kesanggupannya."
(Q.S. Al-Baqarah 2:286)

Kupersembahkan khusus untuk ibu dan ayah. Ibu, Ayah, terimakasih telah mendoaakanku siang dan malam. Rasa bahagia dan banggamu adalah hal yang selalu kuharapkan.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur penulis panjatkan kehadirat Allah SWT, atas segala karunia dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Proposal Tugas Akhir ini dengan judul "Klasifikasi *Android Malware* Menggunakan Algritma *Principal Component Analysis* (PCA) dan *Random Forest*".

Dalam laporan ini penulis menjelaskan mengenai penerapan metode *Principle Component Analysis* (PCA) dan penerapan algoritma *Random Forest* untuk klasifikasi *malware* pada *Android*. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan bagi yang tertarik meneliti tentang *Android malware* serta penerapan *dimensionality reduction* dan klasifikasi *malware* dan *benign*.

Pada penyusunan proposal tugas akhir ini, tidak terlepas dari bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur dan terima kasih kepada yang terhormat :

1. Allah SWT, yang telah memberikan kemudahan, kesehatan, serta kesempatan dalam pelaksanaan pembuatan Tugas Akhir ini.
2. Ibu, Ayah, Nenek serta Adik- adikku Nabilah Nurulhidayah dan Rizki Sefia Nurjannah tercinta yang telah memberikan dukungan dan nasehat-nasehat serta motivasi selama ini. Terima kasih atas dukungan baik berupa moral, material, maupun spiritual.
3. Bapak Jaidan Jauhari, S.Pd., M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Kompuer Fakutas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Deris Stiawan, M.T., PH.D. selaku Pembimbing Tugas Akhir Penulis dan Pembimbing Akademik di Jurusan Sistem Komputer. Terima kasih karena telah meluangkan waktunya untuk membimbing penulis

dalam menyelesaikan tugas akhir ini serta telah memberikan bimbingan dan nasehat selama perkuliahan.

6. Seluruh Dosen Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya.
7. Sahabat-Sahabatku tersayang Arfattustary Noorfizir, Siti Pebsya Roisatun Sholihah, Ulviyana dan Ria Siti Juairiah.
8. Saudara-saudari seperjuanganku yang selalu memberikan semangat motivasi dan bantuan selama hidup diperantauan Sintiya Adelah, Jaya Saputri, Fiddiya Wati dan Desman Aryan.
9. Teman-teman seperjuangan angkatan 2015 dan anak-anak SK15C khususnya yang selalu membersamai selama perkuliahan ini.
10. Serta semua pihak yang telah membantu baik moril maupun materil yang tidak dapat disebutkan satu persatu dalam penyelesaian tugas akhir ini. Terima kasih banyak semuanya.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penulisan Tugas Akhir ini, baik dari materi maupun teknik penyajiannya, mengingat kurangnya pengetahuan dan pengalaman penulis. Untuk itu, penulis mengharapkan adanya kritik dan saran yang membangun agar dapat memperbaiki kekurangan-kekungan tersebut kedepannya nanti.

Akhir kata dengan segala keterbatasan, penulis berharap semoga penulisan Tugas Akhir ini dapat menjadi tambahan wawasan dan ilmu pengetahuan bagi mahasiswa yang memerlukan khususnya mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu pembelajaran.

Indralaya, Desember 2020

Penulis

KLASIFIKASI *ANDROID MALWARE* MENGGUNAKAN ALGORITMA *PRINCIPAL COMPONENT ANALYSIS* (PCA) DAN *RANDOM FOREST*

Dyah Citra Soraya (09011281520107)
Jurusan Sistem Komputer,Fakultas Ilmu Komputer,Universitas Sriwijaya
Email : dyahcitrasoraya@gmail.com

Abstrak

Semakin banyak pihak yang dirugikan karena saat ini malware dapat menjangkit hampir pada seluruh sistem operasi. Salah satu jenis Android malware adalah MazarBot. Mazarbot sangat berbahaya karena apabila sudah ter-install pada perangkat ia dapat mengakses, mengintai dan mengontrol perangkat secara diam-diam dari jarak jauh. Dengan begitu si penyerang dapat memanipulasi serta melakukan apapun yang diinginkan karena mendapatkan akses penuh diperangkat korban. Metode *Random Forest* dapat diterapkan dalam mengklasifikasikan Android Malware. Dimana klasifikasi *Android Malware* berfokus pada *malware Mazarbot* dan *Benign* dengan menggunakan dataset yang bernama CICAndMal2017. Selain itu, metode *Principal Component Analysis* (PCA) juga diterapkan pada penelitian ini fungsinya sebagai *dimensionality redution* yaitu untuk mengurangi jumlah dimensi data yang tinggi menjadi dimensi data yang lebih rendah. Hasil akurasi yang didapat dengan menerapkan metode *Random Forest* adalah 92.06%. Sedangkan untuk akurasi mengunakan gabungan metode *Random Forest* dengan PCA adalah sebesar 82 %.

Kata Kunci : *Android*, *malware*, klasifikasi, *Random Forest*, *Principal Component Analysis* (PCA).

ANDROID MALWARE CLASSIFICATION USING PRINCIPAL COMPONENT ANALYSIS (PCA) AND RANDOM FOREST ALGORITHM

Dyah Citra Soraya (09011281520107)
Jurusan Sistem Komputer,Fakultas Ilmu Komputer,Universitas Sriwijaya
Email : dyahcitrasoraya@gmail.com

Abstrak

More and more parties are harmed because currently malware can infect almost all operating systems. One type of Android malware is MazarBot. Mazarbot is very dangerous because when it is installed on a device it can access, spy on and control the device secretly remotely. That way the attacker can manipulate and do whatever he wants because he gets full access to the victim's device. The Random Forest method can be applied in classifying Android Malware. Where the Android Malware classification focuses on Mazarbot and Benign malware using a dataset called CICAndMal2017. In addition, the Principal Component Analysis (PCA) method is also applied in this study, its function is to reduce the number of high data dimensions to lower data dimensions. The accuracy results obtained by applying the Random Forest method is 92.06%. Meanwhile, the accuracy of using a combination of the Random Forest method with PCA is 82%.

Kata Kunci : *Android, malware,* klasifikasi, *Random Forest*, *Principal Component Analysis* (PCA).

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

PENDAHULUAN

1.1 Latar Belakang

Malware berasal dari kata *malicious*, merupakan file berbahaya yang diciptakan untuk menyelinap ke sistem operasi. *Malwar*e juga disebut file atau *software* perusak karena dapat merusak dan juga mencuri file-file penting yang ada pada sistem. Semakin banyak pihak yang dirugikan karena saat ini *malware* dapat menjangkit hampir pada seluruh sistem operasi [1]. Salah satu jenis *Android malware* adalah *MazarBot*. *Mazarbot* sangat berbahaya karena apabila sudah ter-*install* pada perangkat ia dapat mengakses, mengintai dan mengontrol perangkat secara diam-diam dari jarak jauh. Dengan begitu si penyerang dapat memanipulasi serta melakukan apapun yang diinginkan karena mendapatkan akses penuh diperangkat korban dan tentu saja hal itu sangatlah merugikan [2].

Pada penelitian sebelumnya [3], telah dilakukan Klasifikasi *Android Malware* yang menggunakan beberapa metode yakni *Random Forest, Decision Tree (DT)* dan *K-Nears Nighbor (KNN).* Akan tetapi *dataset* yang digunakan sudah kuno. Dari penelitian tersebut didapatlah hasil terbaik yaitu dengan menggunakan algoritma *Random Forest*. Nilai *recall* yang didapat adalah 88.30%, sedangkan untuk nilai presisi-nya adalah 85.80%.

Dengan mengacu pada penelitian sebelumnya untuk mendapatkan hasil performasi yang lebih baik, maka penulis akan mengimplementasikan

algoritma *Random Forest* untuk mengklasifikasikan *Android Malware.* Dimana klasifikasi *Android Malware* berfokus pada *malware Mazarbot* dan *Benign* dengan menggunakan *dataset* baru yang diusulkan oleh penelitian sebelumnya [3] yaitu *CICAndMal2017*. Selain itu, metode *Principal Component Analysis (PCA)* juga akan diterapkan pada penelitian ini yang berperan sebagai *dimensionality redution* karena *PCA* dapat mengurangi jumlah dimensi data yang tinggi menjadi dimensi data yang lebihmrendah sehingga akan meminimalisir resiko kehilangan informasi.

1.2 Tujuan Penelitian

1. Melakukan klasifikasi *malware mazarbot* dan *benign* menggunakan algoritma *Random Forest*.
2. Menerapkan algoritma *Principal Component Analysis* (PCA) yang digunakan sebagai *dimensionality reduction* pada dataset *CICAndMal2017*.
3. Menganalisa hasil klasifikasi menggunakan algoritma *Random Forest* tanpa diterapkannya *dimensionality reduction.*
4. Menganalisa hasil klasifikasi menggunakan algoritma *Random Forest* yang telah diterapkan algoritma *Principal Component Analysis* (PCA) yang berperan sebagai *dimensionality reduction*.

1.3 Manfaat Penelitian

Hasil dari penelitian ini dapat dijadikan sebagai landasan dalam pengembangan klasifikasi *Android Malware* secara lebih lanjut. Selain itu, manfaat lain dari penelitian ini adalah sebagai berikut :

1. Dapat mengklasifikasikan data yang merupakan *malware mazarbot* dan *benign.*
2. Hasil dari penelitian ini dapat menjadi referensi untuk meningkatkan nilai recall dan presisi yang lebih baik dalam klasifikasi *Android Malware* menggunakan *Random Forest* dan *PCA.*
3. Dapat mempelajari proses dalam klasifikasi *Android malware.*

1.4 Rumusan Masalah

Rumusan masalah yang akan diambil dari penelitian ini adalah bagaimana hasil klasifikasi dari penerapan metode *Principal Component Analysis* dan algoritma *Random Forest* sehingga ditemukan akurasi yang lebih baik dari penelitian sebelumnya.

1.5 Batasan Masalah

Terdapat beberapa batasan masalah yang dirancang dalam tugas akhir ini :

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

2. Klasfikasi *SMS Malware* yang dilakukan adalah secara *binary* (*mazarbot* dan *benign).*
3. Nilai yang diukur yaitu *Accuracy, Recall, Precision, FPR* dan *OOB-Error*.

1.6 Metodelogi Penelitian

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

1. Study Pustaka (Literatur)

   Tahap ini ialah tahapan mencari referensi atau literatur pada *keyword* yang diangkat dari judul yang bertujuan untuk menunjang pada penelitian yang dilakukan.

2. Konsultasi

   Pada tahap ini, peneliti melakukan konsultasi kepada orang-orang yang dianggap memiliki pengetahuan dan wawasan terhadap permasalahan yang ditemui saat pembuatan Tugas Akhir.

3. Pengumpulan Data

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

4. Pengolahan Data

   Pada tahap ini, dilakukan pengolahan data dengan menerapkan algoritma *Principal Component Analysis* sebagai *dimensionality reduction* dan *Random Forest* untuk tahap klasifikasi.

5. Analisa

Pada tahap ini, dilakukan pengambilan data dan menganalisis data yang telah dilakukan pengolahan.

6. Kesimpulan dan Saran

Tahap ini dilakukan penarikan kesimpulan dari analisa dan studi literatur serta saran untuk peulis selanjutnya jika akan dijadikan bahan referensi.

1.7 Sistematika Penulisan

Adapun sistematis penulisan laporan pada tugas akhir ini adalah sebagai berikut :

1. **BAB I PENDAHULUAN**

   Pada bab ini akan dijelaskan mengenai latar belakang, tujuan, manfaat, rumusan masalah, batasan masalah, metodologi penelitian, serta sistematikan penulisan yang akan digunakan dalam laporan penelitian ini.

2. **BAB II TINJAUAN PUSTAKA**

   Bab ini berisikan *literature review* tentang teori-teori yang berkaitan dengan masalah *malware* dengan metode *PCA* dan *Random Forest*.

3. **BAB III METODOLOGI**

   Bab ini akan menjelaskan mengenai uraian dari kerangka kerja, perancangan sistem, langkah kerja dan metodologi yang akan dilakukan dalam proses penelitian ini.

4. BAB IV ANALISA DAN PEMBAHASAN

Bab ini berisikan proses, hasil dan analisa pengujian dari penelitian mengenai klasifikasi *Android SMS Malware* menggunakan *Radom Forest* dan menerapakan algoritma *Principal Component Analysis (PCA)* sebagai *dimensionality reduction.*

5. BAB V KESIMPULAN DAN TINDAK LANJUT

Bab ini mengemukakan kesimpulan dari hasil yang diperoleh dan tindak lanjut untuk penelitian berikutnya.

DAFTAR PUSTAKA

[1] R. Novrianda, Y. N. Kunang, and P. . Shaksono, “Analisis Forensik Pada Platform Android,” *Konf. Nas. ilmu Komput.*, pp. 141–148, 2014.

[2] A. Zaharia, “Security Alert: Mazar BOT Spotted in Active Attacks – the Android Malware That Can Erase Your Phone,” 2016.

[3] 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,” *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2018-Octob, no. Cic, pp. 1–7, 2018.

[4] A. Ahmad, “Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning,” *J. Teknol. Indones.*, no. October, p. 3, 2017.

[5] V. G. Biju, “Kappa and Accuracy Evaluations of Machine Learning Classifiers,” pp. 20–23, 2017.

[6] J. W. G. Putra, “Pengenalan Konsep Pembelajaran dan Deep Learning,” pp. 1–235, 2019.

[7] X. Wang and U. Kingdom, “ALGORITHM BASED ON CO-LEARNING,” no. August, pp. 13–16, 2006.

[8] D. Ucci, L. Aniello, and R. Baldoni, “Survey of machine learning techniques for malware analysis,” *Comput. Secur.*, vol. 81, pp. 123–147, 2019.

[9] K. O. Elish, Yao Danfeng Daphne, and G. R. Barbara, “On the Need of

Precise Inter-App ICC Classification for Detecting Android Malware Collusions," *Proc. Secur. Priv. Work.*, pp. 116–127, 2015.

[10] Khairunnisa and Sutarti, "Perancangan Dan Analisis Keamanan Jaringan Nirkabel Dari Serangan Ddos (Distributed Denial of Service) Berbasis Honeypot," *J. PROSISKO*, vol. 4, no. 2, p. 8, 2017.

[11] V. Kouliaridis, K. Barmpatsalou, G. Kambourakis, and G. Wang, "Mal-warehouse: A data collection-as-a-service of mobile malware behavioral patterns," *Proc. - 2018 IEEE SmartWorld, Ubiquitous Intell. Comput. Adv. Trust. Comput. Scalable Comput. Commun. Cloud Big Data Comput. Internet People Smart City Innov. SmartWorld/UIC/ATC/ScalCom/CBDCo*, no. October, pp. 1503–1508, 2018.

[12] G. Rahayu and M. Mustakim, "Principal Component Analysis Untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru Di Indonesia," *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 201–208, 2017.

[13] M. E-nose, T. W. Widodo, and D. Lelono, "Klasifikasi Teh Hijau dan Teh Hitam Tambi-Pagilaran dengan Metode Principal Component Analysis (PCA)," vol. 8, no. 1, pp. 61–72, 2018.

[14] Q. Wang, Q. Gao, X. Gao, and F. Nie, "Angle principal component analysis," *IJCAI Int. Jt. Conf. Artif. Intell.*, no. August 2018, pp. 2936–2942, 2017.

[15] M. Morchid, R. Dufour, P. M. Bousquet, G. Linarès, and J. M. Torres-

Moreno, “Feature selection using Principal Component Analysis for massive retweet detection,” *Pattern Recognit. Lett.*, vol. 49, pp. 33–39, 2014.

[16] L. E. O. Breiman, “Random Forests,” *Kluwer Acad. Publ.*, pp. 5–32, 2015.

[17] N. S. Intizhami, “Warfare Simulation : Predicting Battleship Winner Using Random Forest,” *2019 IEEE Int. Conf. Commun. Networks Satell.*, pp. 30–34, 2019.

[18] D. Made, S. Arsa, A. Agung, and N. Hary, “VGG16 in Batik Classification based on Random,” *2019 Int. Conf. Inf. Manag. Technol.*, vol. 1, no. August, pp. 295–299, 2019.

[19] M. S. Alam and S. T. Vuong, “Random Forest Classification for Detecting Android Malware,” pp. 663–669, 2013.

[20] L. U. Principal, “Tyang Luhtu,” vol. di, no. v, 2013.