KLASIFIKASI MALWARE ANDROID RANSOMWARE FAMILY DENGAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SUPPORT VECTOR MACHINE (SVM)

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Strata 1

OLEH:

MUHAMMAD RIZALLUL HAKIM
09011381722085

JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN

KLASIFIKASI MALWARE ANDROID RANSOMWARE FAMILY DENGAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SUPPORT VECTOR MACHINE (SVM)

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Strata 1

OLEH:

MUHAMMAD RIZALLUL HAKIM
09011381722085

Palembang, 22 Desember 2021

Pembimbing I	Pembimbing II
Ahmad Heryanto, S.Kom., M.T.	Ahmad Fali Oklilas, M.T.
NIP. 198701222015041002	NIP. 197210151999031001

Mengetahui,
Ketua Jurusan Sistem Komputer

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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada:

Hari : Jumat

Tanggal : 19 November 2021

Tim Penguji:

1. **Ketua Sidang** : **Dr. Ir. H. Sukemi, M.T.**
2. **Sekretaris Sidang** : **Iman Saladin B. Azhar, S.Kom., M.MSI.**
3. **Penguji Sidang** : **Deris Stiawan, M.T., Ph.D.**
4. **Pembimbing I** : **Ahmad Heryanto, M.T.**
5. **Pembimibing II** : **Ahmad Fali Oklilas, M.T.**

Mengetahui, 22/12/21
Ketua Jurusan Sistem Komputer

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

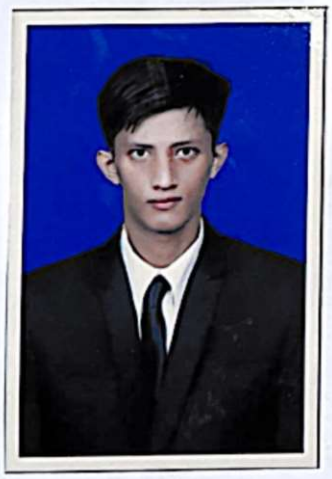
HALAMAN PERNYATAAN

Yang bertanda yangan dibawah ini:

Nama	: Muhammad Rizallul Hakim
NIM	: 09011381722085
Judul	: Klasifikasi Malware Android Ransomware Family Dengan Metode Principal Component Analysis (PCA) Dan Support Vector Machine (SVM)

Hasil pengecekan *Software iThentivate/Turnitin* : 17%

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.



Palembang, November 2021

Muhammad Rizallul Hakim
09011381722085

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 Tugas Akhir ini dengan judul "Klasifikasi Malware Android Ransomware Family Dengan Metode Principal Component Analysis(PCA) Dan Support Vector Machine(SVM)".

Dalam laporan ini penulis menjelaskan mengenai penerapan metode Principle Component Analysis(PCA) dan penerapan algoritma Support Vector Machine(SVM) 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 reduksi dimensi dan klasifikasi *malware* dan *benign.*

Pada penyusunan 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, Adik, serta Keluarga Besar saya yang telah memberikan dukungan dan 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 Ahmad Heryanto, S.Kom., M.T. dan Bapak Ahmad Fali Oklilas, M.T. selaku Pembimbing Tugas Akhir Penulis dan Bapak Dr. Ir. Bambang Tutuko, M.T. selaku 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 Kedal Squad dan yang berada di Bekasi, Cibitung, Tambun serta Palembang yang selalu memberikan semangat, motivasi dan bantuan.
8. Teman-teman seperjuangan angkatan 2017 khususnya yang selalu membersamai selama perkuliahan ini.
9. 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-kekurangan 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.

Palembang, November 2021

Penullis

KLASIFIKASI MALWARE ANDROID RANSOMWARE FAMILY DENGAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SUPPORT VECTOR MACHINE (SVM)

MUHAMMAD RIZALLUL HAKIM (09011381722085)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Univesitas Sriwijaya

Email: rizallul.hakim300@gmail.com

ABSTRAK

Malware adalah program berbahaya yang bertujuan untuk merusak atau mencuri data pribadi dari sistem. Salah satu jenis serangan malware yang digunakan oleh attacker untuk menyerang sistem operasi pada perangkat mobile yaitu jenis malware ransomware. Ransomware ini memiliki banyak varian Salah satunya adalah wannalocker. Wannalocker menyerang platform android dengan cara serupa, menyamar sebagai plugin game, dengan mengelabui korban agar mengunduh dan menjalankannya[6]. Metode Support Vector Machine (SVM) dapat diimplementasikan untuk mengklasifikasikan android malware. Klasifikasi android malware yang berfokus pada malware wannalocker dan benign dengan menggunakan dataset CICAndMal2017. Selain itu, ada metode Principal Component Analysis (PCA) yang berfungsi sebagai reduksi dimensi data pada penelitian ini. Hasil akurasi terbaik yang didapat dengan menerapkan metode SVM dan PCA sebagai reduksi dimensi data ini adalah sebesar 95,02%.

Kata Kunci *: Malware, Android, klasifikasi, Support Vector Machine, Principal Component Analysis*

KLASIFIKASI MALWARE ANDROID RANSOMWARE FAMILY DENGAN METODE PRINCIPAL COMPONENT ANALYSIS (PCA) DAN SUPPORT VECTOR MACHINE (SVM)

MUHAMMAD RIZALLUL HAKIM (09011381722085)

Departement of Computer Engineering, Faculty of Computer Science,

Sriwijaya University

Email: rizallul.hakim300@gmail.com

ABSTRACT

Malware is a malicious program whose purpose is to damage or steal personal data from the system. One type of malware attack used by attackers to attack operating systems on mobile devices is ransomware malware. This ransomware has many variants, one of which is wannalocker. Wannalocker attacks the android platform in a similar way, posing as a game plugin, by tricking victims into downloading and running it[6]. The Support Vector Machine (SVM) method can be implemented to classify android malware. Classification of android malware that focuses on wannalocker and benign malware using the CICAndMal2017 dataset. In addition, there is a Principal Component Analysis (PCA) method that works as a data reduction dimension in this study. The best accuracy results obtained by applying the SVM and PCA methods as dimensional reduction data are 95.02%.

Keywords *: Malware, Android, classification, Support Vector Machine, Principal Component Analysis*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Latar Belakang

Android merupakan sistem operasi berbasis *linux* yang dirancang perangkat *mobile* salah satunya seperti smartphone[1]. Sistem operasi *android* ini telah menjadi sistem operasi *mobile* yang sangat populer[2], Oleh karena itu *android* juga menjadi target serangan oleh *attacker* dengan memanfaatkan *malware*.

Malware adalah suatu program berbahaya yang memiliki tujuan untuk merusak maupun mencuri data pribadi dari sistem. *Malware* juga memiliki beberapa kategori dalam beberapa jenis, seperti *virus, trojan, worm, rootkit, spyware, backdoor, botnet, adware, ransomware* dan sebagainya sesuai dengan perilaku dan penyebaranya[3]. Salah satu jenis serangan malware yang digunakan oleh attacker untuk menyerang sistem operasi pada perangkat mobile yaitu jenis malware ransomware.

Ransomware telah menjadi ancaman global yang sangat serius[4]. Cara kerja *Ransomware* yaitu memblokir akses pengguna dengan cara mengenkripsinya dan meminta tebusan untuk mendapatkan kunci dekripsi[5]. Selain itu juga mengancam untuk mempublikasikan ataupun menghapus data apabila pengguna tidak membayar tebusan.

Ransomware ini memiliki banyak varian Salah satu *family* dari *malware ransomware* pada *android* adalah *wannalocker*. *Wannalocker* menyerang *platform android* dengan cara serupa, menyamar sebagai *plugin game*, dengan mengelabui korban agar mengunduh dan menjalankannya[6]. *Pseudo-Plugin* ini menyandikan data di penyimpanan eksternal dan mengenkripsi file menggunakan algoritma *Advanced Encryption Standard (AES)*[6].

Pada penelitian sebelumnya[7][8][9], telah dilakukan klasifikasi malware android dengan menggunakan beberapa metode seperti *Random Forest*, *Decision Tree*d, dan *K-Nearest Neighbors*. Akan tetapi dataset yang digunakan sudah lawas. Pada algoritma tersebut memiliki kelemahan, seperti pada Random Forest memiliki kelemahan dalam hal kestabilan akurasi yang dihasilkan, *K-Nearest Neighbors* Sensitif terhadap data pencilan (*outlier*), sedangkan pada *Decision Tree* Kesulitan dalam merancang pohon keputusan yang optimal.

Mengacu dari penelitian sebelumnya, penulis akan menggunakan dataset yang baru, yaitu Dataset *CICAndMal2017*. Mengimplementasikan metode *Support Vector Machine (SVM)* untuk klasifikasi *Android Malware Ransomware Wannalocker* dan *Benign*. *SVM* merupakan suatu algoritma atau metode klasifikasi dalam analisis statis maupun dinamis[10]. *SVM* memiliki konsep secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah kelas dengan memaksimalkan jarak antar kelas, Dengan konsep ini *SVM* dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang[11]. Penulis juga akan menerapkan algoritma *Principal component analysis (PCA)* sebagai reduksi dimensi pada dataset, *PCA* ini dapat mereduksi jumlah dimensi data yang tinggi menjadi dimensi data yang lebih rendah.

1.2. Tujuan

Penelitian ini memiliki tujuan, antara lain:

1. Melakukan klasifikasi *malware ransomware wannalocker* dan *Benign* menggunakan algoritma *Support Vector Machine (SVM).*
2. Mengimplementasikan algoritma *Principal component analysis (PCA)* sebagai reduksi dimensi pada dataset.

1.3. Manfaat

Penelitian ini terdapat manfaat, antara lain yaitu:

1. Dapat mengklasifikasi data yang merupakan *Malware Ransomware Wannalocker* dan *Benign*.
2. Mengetahui performa algoritma *Support Vector Machine (SVM)* yang didukung dengan algoritma *Principal component analysis (PCA).*

1.4. Rumusan Masalah

Rumusan masalah yang diambil dari tugas akhir ini adalah:

1. Bagaimana mengklasifikasi data *Malware Ransomware Wannalocker* dengan data *Benign.*
2. Bagaimana menerapkan algoritma *Principal component analysis (PCA)* dalam mendukung algoritma *Support Vector Machine (SVM)* dan mereduksi dimensi data.

1.5. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah:

1. Dataset yang digunakan pada penelitian ini menggunakan *Dataset CICAndMal2017* yang berasal dari *Canadian Institute for Cybersecurity (CIC).*
2. Tidak membahas mengenai deteksi.
3. Klasifikasi yang dilakukan adalah secara *binary*, yaitu *Ransomware Wannalocker* dan *Benign*.

1.6. Metodologi Penelitian

Penelitian ini akan melewati beberapa tahapan sebagai berikut:

1. Tahap Pertama (Studi Pustaka/Literatur)

   Tahap pertama ini dilakukan setelah masalah yang akan dibahas sudah layak untuk diangkat sebagai penelitian, dengan membaca literatur dan merujuk kata kunci yang diangkat dari judul, Sehingga dapat menunjang penulisan laporan Tugas Akhir.

2. Tahap Kedua (Persiapan Data)

   Pada tahap ini dilakukan konversi Data *PCAP* yang didapat dari Dataset *CICAndMal2017* menjadi data *Comma-separated values (CSV)*, Menggunakan Software *CICFlowMeter-4.0.*

3. Tahap Ketiga (Pembersihan Dan Reduksi Data)

   Tahap ini ialah melakukan *Pre-Processing* atau pembersihan pada data dan mereduksi dimensi data tersebut dengan algoritma *Principal component analysis (PCA).*

4. Tahap Keempat

   Pada tahap ini dilakukan klasifikasi antara data *Malware Ransomware Wannalocker* dengan data *Benign* menggunakan algoritma *Support Vector Machine (SVM).*

5. Tahap Kelima

   Pada tahap ini akan ditentukan kesimpulan dari hasil klasifikasi dan studi literatur serta saran untuk penulis selanjutnya yang akan dijadikan untuk bahan acuan.

1.7. Sistematika Penulisan

Dalam proses penyusunan laporan tugas akhir ini, penulis menerapkan sistematika penulisan agar memudahkan dalam mmahami isi dari tiap-tiap bab yang disusun dalam laporan tugas akhir ini. Mengenai sistematika penulisan laporan tugas akhir sebagai berikut:

BAB I PENDAHULUAN

Bab ini akan dijelaskan mengenai latar belakang, tujuan dan manfaat, perumusan masalah, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Bab ini mencakup studi literatur, kerangka teori dan kerangka berfikir yang terkait dengan masalah penelitian ini.

BAB III METODOLOGI

Bab ini menjelaskan tentang langkah-langkah mengenai kerangka kerja, langkah kerja, metodologi dan skenario pengujian.

BAB IV PENGUJIAN DAN ANALISA

Bab ini menjelaskan tentang hasil pengujian dan menganalisis terhadap hasil penelitian.

BAB V KESIMPULAN

Bab ini akan menyajikan kesimpulan yang diambil dari data penelitian yang dilakukan, serta saran untuk pengembangan lebih lanjut dari penelitian ini.

DAFTAR PUSTAKA

[1] P. G. Costa, "Android Malware Detection Using Network Behavior Analysis And Machine Learning Classifiers," no. March, pp. 565–575, 2017.

[2] 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.

[3] N. Udayakumar, T. Subbulakshmi, A. Mishra, S. Mishra, and P. Jain, "Malware category prediction using KNN and SVM classifiers," *Int. J. Mech. Eng. Technol.*, vol. 10, no. 2, pp. 787–797, 2019.

[4] G. Abdulsalamya'u, G. K. Job, S. M. Waziri, B. Jaafar, N. A. Sabongari, and I. Z. Yakubu, "Deep Learning for Detecting Ransomware in Edge Computing Devices Based on Autoencoder Classifier," *4th Int. Conf. Electr. Electron. Commun. Comput. Technol. Optim. Tech. ICEECCOT 2019*, pp. 240–243, 2019.

[5] B. M. Khammas, "Ransomware Detection using Random Forest Technique," *ICT Express*, vol. 6, no. 4, pp. 325–331, 2020.

[6] J. W. Hu, Y. Zhang, and Y. P. Cui, "Research on Android Ransomware Protection Technology," *J. Phys. Conf. Ser.*, vol. 1584, no. 1, 2020.

[7] H. Zhang, S. Luo, Y. Zhang, and L. Pan, "An Efficient Android Malware Detection System Based on Method-Level Behavioral Semantic Analysis," *IEEE Access*, vol. 7, pp. 69246–69256, 2019.

[8] 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.

[9] L. Massarelli, L. Aniello, C. Ciccotelli, L. Querzoni, D. Ucci, and R. Baldoni, "AndroDFA: Android malware classification based on resource consumption," *Inf.*, vol. 11, no. 6, pp. 1–20, 2020.

[10] Y. F. Lu, C. F. Kuo, H. Y. Chen, C. W. Chen, and S. C. Chou, "A SVM-Based Malware Detection Mechanism for Android Devices," *2018 Int. Conf. Syst. Sci. Eng. ICSSE 2018*, pp. 1–6, 2018.

[11] K. A. Rokhman, B. Berlilana, and P. Arsi, "Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online," *J. Inf. Syst. Manag.*, vol. 3, no. 1, pp. 1–7, 2021.

[12] M. Abuthawabeh and K. Mahmoud, "Enhanced android malware detection and family classification, using conversation-level network traffic features," *Int. Arab J. Inf. Technol.*, vol. 17, no. 4 Special Issue, pp. 607–614, 2020.

[13] S. Herlambang, S. Basuki, D. R. Akbi, and Z. Sari, "Deteksi Malware Android

Berdasarkan System Call Menggunakan Algortima Support Vector Machine," vol. 5, pp. 157–165, 2015.

[14] V. Rahmayanti *et al.*, "Klasifikasi Malware Family Menggunakan Metode K-Nearest Neighbor," pp. 319–323, 2020.

[15] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed, and M. Xu, "A Survey on Machine Learning Techniques for Cyber Security in the Last Decade," *IEEE Access*, vol. 8, pp. 222310–222354, 2020.

[16] J. Ma and Y. Yuan, "Dimension reduction of image deep feature using PCA," *J. Vis. Commun. Image Represent.*, vol. 63, 2019.

[17] L. Onwuzurike, E. Mariconti, P. Andriotis, E. De Cristofaro, G. Ross, and G. Stringhini, "MaMaDroid: Detecting android malware by building markov chains of behavioral models (extended version)," *arXiv*, vol. 22, no. 2, 2017.

[18] Y.-M. Kwon, J.-J. An, M.-J. Lim, S. Cho, and W.-M. Gal, "Malware Classification Using Simhash Encoding and PCA (MCSP)," *Symmetry (Basel).*, vol. 12, no. 5, p. 830, 2020.

[19] A. Directions, "Principal Component Analysis (PCA) Principal Component Analysis (PCA)," *Statistics (Ber).*, no. June, pp. 1–12, 2007.

[20] S. Prayoginingsih and R. P. Kusumawardani, "Klasifikasi Data Twitter Pelanggan Berdasarkan Kategori myTelkomsel Menggunakan Metode Support Vector Machine (SVM)," *Sisfo*, vol. 07, no. 02, 2018.

[21] A. Tharwat, "Parameter investigation of support vector machine classifier with kernel functions," *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1269–1302, 2019.

[22] F. Karimi, S. Sultana, A. Shirzadi Babakan, and S. Suthaharan, "An enhanced support vector machine model for urban expansion prediction," *Comput. Environ. Urban Syst.*, vol. 75, no. August 2018, pp. 61–75, 2019.

[23] B. Richhariya, M. Tanveer, and A. H. Rashid, "Diagnosis of Alzheimer's disease using universum support vector machine based recursive feature elimination (USVM-RFE)," *Biomed. Signal Process. Control*, vol. 59, 2020.

[24] M. Yang, X. Chen, Y. Luo, and H. Zhang, "An Android Malware Detection Model Based on DT-SVM," *Secur. Commun. Networks*, vol. 2020, 2020.

[25] S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics and Proteomics*, vol. 15, no. 1, pp. 41–51, 2018.

[26] A. M. Puspitasari, D. E. Ratnawati, and A. W. Widodo, "Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 2, pp. 802–810, 2018.

[27] V. K. Chauhan, K. Dahiya, and A. Sharma, "Problem formulations and solvers in linear SVM: a review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 803–855, 2019.