KLASIFIKASI SMS MALWARE PADA PLATFORM ANDROID MENGGUNAKAN METODE RANDOM FOREST

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer

OLEH :

Jonathan Jeremia Valentino Vici Sitohang

09011181823007

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2022

LEMBAR PENGESAHAN

KLASIFIKASI SMS MALWARE PADA PLATFORM ANDROID MENGGUNAKAN METODE RANDOM FOREST

TUGAS AKHIR

Program Studi Sistem Komputer
Jenjang S1

Oleh

Jonathan Jeremia Valentino Vici Sitohang

09011181823007

Indralaya, Juli 2022

Mengetahui,

Ketua Jurusan Sistem Komputer	Pembimbing Tugas Akhir
Dr. Ir. H. Sukemi, M.T.	**Deris Stiawan. M.T.. Ph.D., IPU.**
NIP. 196612032006041001	**NIP. 197806172006041002**

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Rabu

Tanggal : 20 Juli 2022

Tim Penguji :

1. **Ketua Sidang** : **Ahmad Zarkasi, M.T.**
2. **Sekretaris Sidang** : **Adi Hermansyah, M.T.**
3. **Penguji Sidang** : **Ahmad Heryanto, M.T.**
4. **Pembimbing** : **Deris Stiawan, M.T., Ph.D., IPU**

Mengetahui,

Ketua Jurusan Sistem Komputer

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

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Jonathan Jeremia Valentino Vici Sitohang

NIM : 09011181823007

Judul : Klasifikasi *SMS Malware* Pada Platform Android Menggunakan Metode *Random Forest*

Hasil Pengecekan Software iThenticate/Turnitin : 7%

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 tanpa paksaan dari siapapun.

Palembang, Juli 2022

Jonathan Jeremia V.V.S.

NIM. 09011181823007

KATA PENGANTAR

Shalom,

Segala puji dan syukur bagi Tuhan Yesus Kristus, dikarenakan karunia dan anugerah-Nya yang melimpah serta kemurahan dan kasih setia yang besar, sehingga Penulis dapat menyelesaikan penulisan Proposal Tugas Akhir ini yang berjudul **"Klasifikasi SMS Malware pada Platform Android Menggunakan Metode Random Forest"**.

Pada laporan ini Penulis menuliskan bagaimana caranya untuk melakukan klasifikasi pada *malware* SMS yang didapatkan Penulis saat melakukan penelitian dan pengujian data. Penuis berharap agar kedepannya tulisan ini dapat digunakan oleh orang lain dan menjadi bacaan yang menarik pada bidang *malware*.

Pada kesempatan ini, penulis ingin mengucapkan terima kasih kepada berbagai pihak yang telah membantu seperti memberikan ide dan saran dalam pengerjaan penulisan Proposal Tugas Akhir ini. Oleh karena itu Penulis ingin mengucapkan rasa syukur serta mengucapkan terima kasih kepada:

1. Tuhan Yesus Kristus yang telah memberikan karunia dan anugerah-Nya sehingga penulis dapat menyelesaikan penulisan Proposal Tugas Akhir ini dengan baik dan lancar.
2. Orang tua saya yang telah membesarkan dan mendidik saya dengan tegas dan penuh kasih sayang dan selalu mengajarkan saya untuk berbuat hal yang baik. Terima kasih untuk segala motivasi, doa, serta dukungan moril, materil dan spritiual selama ini.
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 Komputer Fakultas Ilmu Komputer Universitas Sriwijaya
5. Bapak Deris Stiawan. S.Kom, M.T, Ph.D. selaku Dosen Pembimbing Tugas Akhir yang telah berkenan memberikan waktunya untuk

membimbing serta memberikan saran terbaik untuk Penulis untuk menyelesaikan Tugas Akhir.

6. Bapak Dr. Ir. Bambang Tutuko, M.T. selaku Pembimbing Akademik Jurusan Sistem Komputer.
7. Dan semua pihak yang telah membantu Penulis dalam penelitian.

Penulis menyadari bahwa laporan ini masih sangat jauh dari kesempurnaan. Oleh karena itu kritik dan saran yang membangun sangatlah diperlukan dan diharapkan oleh Penulis agar menjadi suatu pemikiran baru sehingga Penulis dapat segera memperbaiki sehingga laporan ini dapat menjadi suatu masukan ide dan pemikiran bagi semua pihak dan dapat menjadi bahan bacaan yang menarik bagi yang tertarik dalam penelitian pada bidang klasifikasi *Malware*.

Amin.

Indralaya, Juli 2022

Jonathan Jeremia V.V.S.
NIM. 09011181823007

KLASIFIKASI *SMS MALWARE* PADA PLATFORM ANDROID MENGGUNAKAN METODE *RANDOM FOREST*

Jonathan Jeremia Valentino Vici Sitohang (09011181823007)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwjaya
Email : jonathansitohang2902@gmail.com

ABSTRAK

Android merupakan salah satu platform seluler paling populer di dunia sebagai sistem operasi seluler, masih menghadapi tantangan dan ancaman keamanan. Android adalah malware dengan berbagai varian yang dimasukkan dalam package aplikasi yang memiliki ekstensi APK (*Android Package Kit*) dengan memiliki *permission* untuk SMS. SMS (*Short Messages Service*) merupakan suatu fitur layanan standar yang berada pada *smartphone* kini. Teknologi yang berkembang yakni Android memiliki fitur tersebut yang berarti membuka juga gerbang varian *malware* yang tersedia yang dapat menyerang layanan SMS. SMS-*based Malware* kemudian ditemukan pada tahun 2012. Singkatnya, SMS *Malware* ini dapat melakukan serangan melalui SMS dan dapat melakukan *subscriptions* (langganan) pada suatu aplikasi secara diam-diam. Penelitian ini menggunakan dataset yang disediakan oleh *Canadian Institute of Cybersecurity* yang memiliki dataset *SMS Malware BeanBot*. Hasil dari klasifikasi menggunakan normalisasi *MinMaxScaler* serta seleksi fitur *Chi-Square* yang ditujukan untuk mendapatkan fitur terbaik pada dataset, yang kemudian dilakukan klasifikasi menggunakan algoritma *Random Forest*. Penelitian ini menggunakan berbagai variasi split data dengan *test size* 0.2, 0.3, dan 0.4. Hasil klasifikasi dari penelitian ini menghasilkan bahwa seleksi fitur *Chi-Square* dengan 30 fitur pilihan dengan *test size* 0.2 mendapatkan hasil terbaik yakni akurasi sebesar 89.53%, *Recall* sebesar 93,49%, *Precision* sebesar 87.39%, nilai *False Positive Rate* dan nilai *error* terbaik berada pada *test size* 0.4 dengan nilai *False Positive Rate* sebesar 9.67% dan *Error* sebesar 10.46%.

Kata Kunci : *Short Message Service, SMS Malware Classification, Random Forest, Machine Learning*

CLASSIFICATION OF SMS MALWARE ON ANDROID PLATFORM USING RANDOM FOREST METHOD

Jonathan Jeremia Valentino Vici Sitohang (09011181823007)
Department of Computer Engineering , Faculty of Computer Science, Sriwijaya University
Email: jonathansitohang2902@gmail.com

ABSTRACT

Android is one of the most popular mobile platforms in the world as a mobile operating system, still facing security challenges and threats. Android is malware with various variants that are included in application packages that have APK (Android Package Kit) extensions with permission for SMS. SMS (Short Messages Service) is a standard service feature on today's smartphones. The growing technology of Android has this feature, which means that it also opens the gate for available malware variants that can attack SMS services. SMS-based Malware was then discovered in 2012. In short, SMS Malware can perform attacks via SMS and can perform subscriptions to an application silently. This research uses a dataset provided by the Canadian Institute of Cybersecurity which has the BeanBot SMS Malware dataset. The results of the classification use MinMaxScaler normalization and Chi-Square feature selection aimed at getting the best features on the dataset, which are then classified using the Random Forest algorithm. This research uses various variations of split data with test sizes 0.2, 0.3, and 0.4. The classification results of this study show that the Chi-Square feature selection with 30 selected features with a test size of 0.2 gets the best results, namely accuracy of 89.53%, Recall of 93.49%, Precision of 87.39%, False Positive Rate value and the best error value is at test size 0.4 with a False Positive Rate value of 9.67% and Error of 10.46%.

Keywords : *Short Message Service, SMS Malware Classification, Random Forest, Machine Learning*

DAFTAR ISI

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR TABEL

Halaman

BAB I
PENDAHULUAN

1.1 Latar Belakang

Android merupakan salah satu platform seluler paling populer di dunia sebagai sistem operasi seluler [1], masih menghadapi tantangan dan ancaman keamanan dari aplikasi yang mencurigakan, alasannya karena sistem operasi android bersifat open source, desain izin (*permission*) yang masih disempurnakan hingga sekarang, serta tidak adanya sertifikasi penuh dalam mempublikasi aplikasi ke sistem operasi tersebut [2]. Salah satu ancaman yang dapat menyerang Android adalah malware dengan berbagai varian yang dimasukkan dalam package aplikasi yang memiliki ekstensi APK (*Android Package Kit*). Saat ini, popularitas perangkat Android menjadikannya target yang diinginkan [3].

Ancaman malware yang ada pada sistem operasi android tersebut dapat dimanfaatkan oleh *hacker* untuk mengeksploitasi fitur pada aplikasi android untuk merusak keamanan dan privasi perangkat, yang menimbulkan ancaman serius terhadap kebocoran data pribadi seperti lokasi pengguna, informasi kontak, akun, foto, dan lain-lain. Selain itu, sebagian besar perangkat Android tidak menggunakan aplikasi deteksi anti-virus atau malware[4].

SMS (*Short Messages Service*) merupakan suatu fitur layanan standar yang berada pada smartphone kini. Teknologi yang berkembang yakni Android memiliki fitur tersebut yang berarti membuka juga gerbang varian malware yang tersedia yang dapat menyerang layanan SMS [5]. Nyatanya, pada tahun 2012 perusahaan antivirus menemukan masalah baru varian malware, yakni dinamakan SMS-based Malware. Singkatnya, *SMS Malware* ini dapat melakukan serangan melalui SMS dan dapat melakukan *subscriptions* (langganan) pada suatu aplikasi secara diam-diam [5].

Pada survei machine learning untuk *malware*[6], disebutkan bahwa SVM, KNN, Random Forest, dan Decision Tree digunakan untuk melakukan klasifikasi pada malware dengan Random Forest yang memiliki performa yang

baik dalam menghitung TPR/FPR[6]. Pada penelitian sebelumnya[1], dilakukan penelitian yang melakukan klasifikasi *malware Android* yang menggunakan metode *Random Forest*, *K-Nearest Neighbors* (KNN), dan *Decision Tree* (DT). Hasil dari penelitian tersebut menunjukan bahwa algoritma *Random Forest* mendapatkan hasil terbaik dari penelitian tersebut dengan nilai *recall* 88.30% dan nilai presisi 85.80% dengan menggunakan dataset lawas atau kuno.

Berdasarkan penjelasan serta penelitian sebelumnya yang akan dijadikan acuan penelitian kali ini, maka Penulis akan melakukan klasifikasi pada dataset baru yang diusulkan dari penelitian sebelumnya yakni *CICAndMal2017*[1] yang berfokus pada *SMS Malware*, sehingga penelitian ini akan diberikan judul "Klasifikasi SMS Malware pada Platform Android Menggunakan Metode Random Forest"

1.2 Perumusan Masalah

Rumusan masalah yang akan diangkat pada penelitian ini adalah bagaimana menerapkan klasifikasi *SMS Malware BeanBot* menggunakan algoritma *Random Forest*. Sehingga ditemukan hasil klasifikasi yang lebih baik dari penelitian sebelumnya.

1.3 Batasan Masalah

Batasan masalah yang terjadi dalam penelitian penelitian ini adalah:

1. Dataset yang digunakan merupakan dataset terbatas yang berasal dari *Canadian Institute for Cybersecurity* (CIC) yakni dataset CICAndMal2017 yang menyasar pada beberapa varian *SMS Malware* saja yakni *malware BeanBot* serta *Benign* file.
2. Klasifikasi yang dilakukan hanya berada pada varian terbatas pada *SMS Malware* yakni *BeanBot* serta *Benign* file menggunakan metode klasifikasi *Random Forest*.
3. Klasifikasi *SMS Malware* dilakukan secara *binary* (*BeanBot* dan *Benign*)

1.4 Tujuan Penelitian

Tujuan dari penelitian ini yakni:

1. Melakukan klasifikasi *malware BeanBot* dan *Benign* menggunakan algoritma *Random Forest*.
2. Menerapakan metode *Random Forest* sebagai seleksi fitur dalam SMS Malware.
3. Membuat analisa untuk hasil klasifikasi menggunakan algoritma *Random Forest*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yakni:

1. Dapat melakukan klasifikasi dataset yang merupakan *malware BeanBot* dan *Benign*.
2. Dapat mengetahui tingkat akurasi algoritma klasifikasi *Random Forest* dalam melakukan klasifikasi *malware*.
3. Dapat mempelajari proses dalam melakukan klasifikasi dataset menggunakan algoritma *Random Forest*.

1.6 Metodologi Penelitian

Metodologi yang dilakukan akan melalui tahap berikut, yakni:

1. Studi Pustaka (*Literature*)

   Pada metode ini, dilakukanlah tahap mencari informasi serta masalah yang sesuai dan revelan yang dapat digunakan dalam penelitian, dengan mencari referensi atau literatur yang sesuai dengan topik yang diteliti seperti di buku, artikel, dan lain sebagainya yang menunjang penelitian.

2. Pengumpulan dan Perancangan Sistem

   Pada tahap ini dibuatlah suatu rancangan sistem untuk melakukan penelitian dan menerapkan metode yang telah ditentukan serta menyiapkan data yang berupa dataset dan perangkat keras hingga perangkat lunak untuk melakukan konfigurasi serta menulis code.

3. Pengujian Data

Pada tahap ini, dilakukan pengujian dengan menggunakan dataset yang berasal dari *Canadian Institute for Cybersecurity* (CIC) yakni dataset CICAndMal2017 pada kategori SMS Malware dan *Benign*.

4. Analisa

Selanjutnya pada tahap ini dilakukan pengujian pada dataset yang telah siap untuk diolah dan diklasifikasi. Serta akan kembali ke tahap ketiga apabila ada kesalahan dalam seleksi fitur atau akurasi yang diberikan tidak sesuai dengan harapan.

5. Kesimpulan dan Saran

Pada tahap ini, analisa dilakukan untuk menganalisa data pengujian untuk menentukan apakah rancangan sistem berjalan baik atau tidak, sehingga memberikan data yang optimal. Kesimpulan dan saran akan dibuat berdasarkan permasalahan yang ada serta hasil dari analisa pengujian.

1.7 Sistematika Penulisan

Sistematis penulisan laporan yang akan dilakukan pada tugas akhir ini mencakup berikut:

1. **BAB I PENDAHULUAN**

   Bab ini akan menjelaskan mengenai latar belakang, tujuan penelitian, manfaat penelitian, rumusan masalah, batasan masalah, metodologi penelitian, serta sistematika penulisan yang digunakan pada penelitian ini.

2. **BAB II TINJAUAN PUSTAKA**

   Bab ini berisikan bacaan literature (*literature review*) untuk menunjang penelitian yang berupa teori-teori yang berkaitan tentang masalah malware.

3. **BAB III METODOLOGI PENELITIAN**

   Bab ini akan menjelaskan bagaimana penelitian ini berkerja, yang berupa uraian mengenai kerangka kerja, perancangan sistem, langkah kerja dan metodologi yang dilakukan pada penelitian ini.

4. **BAB IV HASIL DAN ANALISA**

Bab ini akan menjelaskan mengenai hasil dari pengujian klasifikasi *SMS Malware* menggunakan algoritma *Random Forest*. Serta hasil dari klasifikasi tersebut akan dibuat analisa yang menjelaskan hasil penelitian.

5. **BAB V KESIMPULAN**

Bab ini akan menjelaskan mengenai kesimpulan dari hasil penelitian

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," *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2018-Octob, no. Cic, pp. 1–7, 2018, doi: 10.1109/CCST.2018.8585560.

[2] X. Liu, X. Du, Q. Lei, and K. Liu, "Multifamily classification of android malware with a fuzzy strategy to resist polymorphic familial variants," *IEEE Access*, vol. 8, pp. 156900–156914, 2020, doi: 10.1109/ACCESS.2020.3019282.

[3] H. Chen, J. Su, L. Qiao, and Q. Xin, "Malware collusion attack against SVM: Issues and countermeasures," *Appl. Sci.*, vol. 8, no. 10, 2018, doi: 10.3390/app8101718.

[4] R. Kumar, X. Zhang, W. Wang, R. U. Khan, J. Kumar, and A. Sharif, "A Multimodal Malware Detection Technique for Android IoT Devices Using Various Features," *IEEE Access*, vol. 7, pp. 64411–64430, 2019, doi: 10.1109/ACCESS.2019.2916886.

[5] K. Hamandi, A. Chehab, I. H. Elhajj, and A. Kayssi, "Android SMS malware: Vulnerability and mitigation," *Proc. - 27th Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2013*, pp. 1004–1009, 2013, doi: 10.1109/WAINA.2013.134.

[6] 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, doi: 10.1109/ACCESS.2020.3041951.

[7] L. Taheri, A. F. A. Kadir, and A. H. Lashkari, "Extensible android malware detection and family classification using network-flows and API-calls," *Proc. - Int. Carnahan Conf. Secur. Technol.*, vol. 2019-Octob, no. Cic, 2019, doi: 10.1109/CCST.2019.8888430.

[8] M. K. A. Abuthawabeh and K. W. Mahmoud, "Android malware detection and categorization based on conversation-level network traffic features,"

Proc. - 2019 Int. Arab Conf. Inf. Technol. ACIT 2019, pp. 42–47, 2019, doi: 10.1109/ACIT47987.2019.8991114.

[9] O. N. Elayan and A. M. Mustafa, "Android malware detection using deep learning," *Procedia Comput. Sci.*, vol. 184, no. 2019, pp. 847–852, 2021, doi: 10.1016/j.procs.2021.03.106.

[10] M. Gohari, S. Hashemi, and L. Abdi, "Android Malware Detection and Classification Based on Network Traffic Using Deep Learning," *2021 7th Int. Conf. Web Res. ICWR 2021*, pp. 71–77, 2021, doi: 10.1109/ICWR51868.2021.9443025.

[11] Y. Zhou and X. Jiang, "Dissecting Android malware: Characterization and evolution," *Proc. - IEEE Symp. Secur. Priv.*, no. 4, pp. 95–109, 2012, doi: 10.1109/SP.2012.16.

[12] C. D. Morales-Molina, D. Santamaria-Guerrero, G. Sanchez-Perez, H. Perez-Meana, and A. Hernandez-Suarez, "Methodology for malware classification using a random forest classifier," *2018 IEEE Int. Autumn Meet. Power, Electron. Comput. ROPEC 2018*, no. Ropec, pp. 1–6, 2019, doi: 10.1109/ROPEC.2018.8661441.

[13] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, "Improved Random Forest for Classification," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4012–4024, 2018, doi: 10.1109/TIP.2018.2834830.

[14] S. Shalev-Shwartz and S. Ben-David, *Understanding machine learning: From theory to algorithms*, vol. 9781107057135. 2013. doi: 10.1017/CBO9781107298019.

[15] M. S. Alam and S. T. Vuong, "Random forest classification for detecting android malware," *Proc. - 2013 IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Soc. Comput. GreenCom-iThings-CPSCom 2013*, pp. 663–669, 2013, doi: 10.1109/GreenCom-iThings-CPSCom.2013.122.