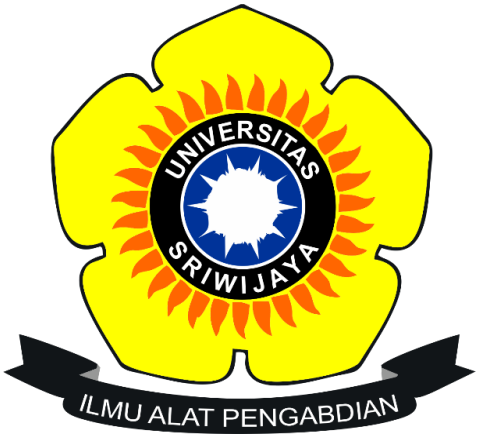
KLASIFIKASI *SMS MALWARE* PADA PLATFORM *ANDROID* DENGAN METODE *SUPPORT VECTOR MACHINE*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer



OLEH :

AGENG RAHARJO

09011381924133

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN

KLASIFIKASI *SMS MALWARE* PADA PLATFORM *ANDROID* DENGAN METODE *SUPPORT VECTOR MACHINE*

TUGAS AKHIR

Program Studi Sistem Komputer
Jenjang S1

Oleh

AGENG RAHARJO

09011381924133

Palembang, 24 Juli 2023

Mengetahui,

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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jumat

Tanggal : 7 Juni 2023

Tim Penguji :

1. **Ketua** : **Huda Ubaya, M.T.**
2. **Sekretaris** : **Iman Saladin B Azhar, M.MSI,**
3. **Penguji** : **Ahmad Heryanto, M.T.**
4. **Pembimbing** : **Deris Stiawan, M.T., Ph.D.**

Mengetahui, 24 Juli 2023

Ketua Jurusan Sistem Komputer

Dr. Ir. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama	:	Ageng Raharjo
NIM	:	09011381924133
Judul	:	Klasifikasi *SMS Malware* pada Platform *Android* Dengan Metode *Support Vector Machine*

Hasil Pengecekan Software Turnitin : 8%

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, 11 Juli 2023

Ageng Raharjo

NIM. 09011381924133

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Segala puji dan syukur atas kehadirat Allah SWT, karena berkat Rahmat dan Karunia-Nya lah sehingga penulis dapat menyelesaikan Proposal Tugas Akhir dengan judul **"Klasifikasi *Sms Malware* Pada Platform *Android* Dengan Metode *Support Vector Machine*"**. Shalawat beriringkan salam senantiasa tercurahkan kepada Nabi Muhammad Shallallaahu 'Alaihi Wasallam yang telah membawa kedamaian dan rahmat untuk semesta alam serta menjadi suri tauladan bagi umatnya.

Proposal ini merupakan salah satu syarat untuk memenuhi sebagian kurikulum dan syarat kelulusan Mata Kuliah Tugas Akhir pada Jurusan Sistem Komputer, Universitas Sriwijaya. Selesainya penulisan Proposal Tugas Akhir ini tidak terlepas dari peran serta semua pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada :

1. Allah Subhanahu Wata'ala yang telah memberikan berkah serta nikmat Kesehatan dan kesempatan kepada penulis untuk menyelesaikan Proposal Tugas Akhir.
2. Kedua Orang Tua dan Keluarga yang selalu mendoakan serta memberikan motivasi dan semangat.
3. Bapak Alm. Jaidan Jauhari, S.Pd., M.T. Selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Julian Supardi, S.Pd., M.T. selaku Wakil Dekan Bidang Akademik di Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Dr. Ir. H. Sukemi., M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
6. Bapak Abdurahman, S.KOM., M. Han. selaku Dosen Pembimbing Akademik.

7. Bapak Deris Stiawan, M.T., Ph.D., IPU., ASEAN ENG. selaku Dosen Pembimbing Tugas Akhir serta Dosen Pembimbing Akademik yang telah berkenan meluangkan waktunya untuk membimbing, memberikan saran dan motivasi serta bimbingan terbaik untuk penulis dalam menyelesaikan Tugas Akhir ini.
8. Mbak Nurul M.Kom selaku dosen Sistem Komputer yang telah memberikan bimbingan dan saran selama penulisan tugas akhir ini.
9. Mbak Sari selaku admin jurusan Sistem Komputer yag telah membantu mengurus administrasi dan berkas selama kuliah ini.
10. Terima kasih, Grup Senja 2.0 ! Terima kasih atas kehadiran kalian setiap malam di Gmeet, menemani saya dalam penulisan skripsi.
11. Terima kasih COMNETS LAB atas bantuan fasilitas dan sarana dalam menyelesaikan tugas akhir.
12. Teman-teman di laboratorium COMNETS dan teman-teman jurusan Sistem Komputer Unggulan 2019.
13. Seluruh pihak yang tidak dapat penulis sebutkan satu persatu, yang telah memberikan semangat serta do'a.

Dalam penyusunan tugas akhir ini, penulis menyadari bahwa masih banyak terdapat kekurangan. Untuk itu, penulis mengharapkan masukan berupa kritik dan saran yang membangun demi kesempuraan di masa depan. Akhir kata penulis berharap, semoga tugas akhir ini dapat menjadi amal ibadah dan bermanfaat bagi orang lain.

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Palembang, 24 Juli 2023

Penulis,

Ageng Raharjo
NIM. 09011381924133

KLASIFIKASI *SMS MALWARE* PADA *PLATFORM ANDROID* DENGAN METODE *SUPPORT VECTOR MACHINE*

Ageng Raharjo (09011381924133)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : agengraharjo2001@gmail.com

ABSTRAK

Android adalah *platform* perangkat lunak seluler yang paling populer di seluruh belahan bumi, unduhan aplikasi pada seluruh dunia pada tahun 2021 mencapai 352,9 milliar, masih menghadapi kesulitan dari keancaman keamanan yang serius, itu disebabkan karena *android* sifatnya *open-source*. *Android* adalah *malware* dengan berbagai varian yang dimasukkan dalam *package* aplikasi yang memiliki ekstensi APK (*Android Package Kit*) dengan memiliki permission untuk SMS. *Short Messages Service* (SMS) merupakan teknologi yang digunakan untuk mengirim pesan teks di *Android* saat ini. SMS dengan pesatnya teknologi saat ini tidak menutup kemungkinan teknologi ini disusupi dengan serangan atau kejahatan *malware. malware* jenis ini dapat diunduh secara otomatis dan tanpa sepengetahuan pengguna melalui pesan singkat atau SMS. Setelah terunduh, *malware* dapat memasang aplikasi-aplikasi jahat lainnya pada perangkat pengguna. Hal ini dapat membahayakan privasi dan keamanan data pribadi pengguna. penelitian ini akan menggnakan dataset yang berasal dari *CICAndMal2017* yang lebih berfokus pada SMS *Malware*, dengan jenis *jifake* dan *fakenotify* menggunakan algoritma *Support Vector Machine*. Hasil klasifikasi dengan kernel RBF memperoleh nilai presisi sebesar 94,67%, *recall* sebesar 94,67%, *F-1 Score* sebesar 94,67% dan akurasi 95% dan *Support Vector Machine* dengan kernel Polynomial memperoleh nilai presisi sebesar 89,67%, *recall* sebesar 90,33%, *F1-Score* sebesar 89,67% dan akurasi sebesar 90%.

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

CLASSIFICATION OF SMS MALWARE ON ANDROID PLATFORM USING SUPPORT VECTOR MACHINE METHOD

Ageng Raharjo (09011381924133)
Computer Engineering Department, Computer Science Faculty, Sriwijaya University
Email : agengraharjo2001@gmail.com

ABSTRACT

Android is the most popular mobile software platform across the globe. The worldwide app downloads reached 352.9 billion in 2021. However, it still faces serious security threats due to its open-source nature. Android is susceptible to various malware variants that are packaged within APK (Android Package Kit) files and have permissions for SMS (Short Message Service). SMS is a technology used for sending text messages on Android devices. With the rapid advancement of technology, SMS is not immune to attacks or malicious activities. This type of malware can be automatically downloaded without the user's knowledge through short messages or SMS. Once downloaded, the malware can install other malicious applications on the user's device, posing a risk to privacy and personal data security. This research utilizes a dataset from CICAndMal2017 that focuses more on SMS malware, specifically the jifake and fakenotify types, using the Support Vector Machine algorithm. The classification results using the RBF kernel achieved a precision of 94.67%, recall of 94.67%, F-1 Score of 94.67%, and an accuracy of 95%. Meanwhile, the Support Vector Machine with the Polynomial kernel obtained a precision of 89.67%, recall of 90.33%, F1-Score of 89.67%, and an accuracy of 90%.

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

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I

PENDAHULUAN

1.1 Latar Belakang

Android adalah platform perangkat lunak seluler yang paling populer di seluruh belahan bumi, unduhan aplikasi pada seluruh dunia pada tahun 2021 mencapai 352,9 milliar[1], masih menghadapi kesulitan dari keancaman keamanan yang serius, itu disebabkan karena *Android* sifatnya *open-source*[2], sistem izin yang kurang sempurna, dan tidak adanya sertifikat penuh untuk publikasi aplikasi.

Malware adalah perangkat lunak (*software)* atau sebuah sistem yang dibuat untuk tujuan menemukan kelemahan bahkan melakukan kerusakan pada sistem komputer atau jaringan komputer secara tidak sah(*permission)*[3]. Ancaman *Malware* [2] pada sistem operasi dapat mengeksploitasi kerentanan dalam sistem Android untuk melakukan berbagai serangan, seperti eskalasi hak istimewa, pengendalian jarak jauh, pengisian biaya finansial, dan pencurian informasi pribadi, sangat mengancam perlindungan privasi, keamanan finansial, bahkan stabilitas sosial.

Short Messages Service(SMS) merupakan teknlogi yang digunakan untuk mengirim pesan teks di Android saat ini. SMS dengan pesatnya teknologi saat ini tidak menutup kemungkinan teklnologi ini disusupi dengan serangan atau kejahatan malware[4]. Pada penelitian ini[5] menemukan *Trojan VDloader* dari *NQ mobile*, *malware* jenis ini dapat diunduh secara otomatis dan tanpa sepengetahuan pengguna melalui pesan singkat atau SMS. Setelah terunduh, *Trojan VDloader* dapat memasang aplikasi-aplikasi jahat lainnya pada perangkat pengguna. Hal ini dapat membahayakan privasi dan keamanan data pribadi pengguna.

Pada penelitian ini[6] algoritma *Support Vector Machine* digunakan untuk mendeteksi serangan *Android Malware* dengan dataset [7]. Kernel Linear mendapatkan nilai recall, precision dan akurasi lebih baik dibandingkan dengan

kernel Sigmoid, Polynomial dan RBF dan dari penelitian ini algoritma yang menggunakan parameter kernel linear mendapatkan nilai recall 99.97%, precision 99.54% dan akurasi 99.75%. pada penelitian lainnya[8] menunjukkan algoritma SVM mendapatkan nilai akurasi lebih baik dibandingkan metode lain seperti *BayesNet, NaiveBayes ,NBTree, CART, Random Tree* dan J48 untuk klasifikasi serangan *Android Malware* dengan hasil akurasi yaitu 90.08%.

Dengan merujuk pada penjelasan dan penelitian sebelumnya sebagai referensi, penelitian ini akan menggnakan dataset yang berasal dari CICAndMal2017[1] yang lebih berfokus pada SMS Malware. Selanjutnya, melakukan klasifikasi menggunakan salah satu metode dalam *machine learning*. Oleh karena itu, penelitian ini akan berjudul "Klasifikasi SMS Malware pada Platform Android Menggunakan Metode Support Vector Machine".

1.2 Rumusan Masalah

Penelitian ini memasukan rumusan masalah sebagai berikut:

1. Bagaimana langkah-langkah persiapan data yang dilakukan untuk mengklasifikasi *SMS Malware*?
2. Bagaimana langkah-langkah untuk melakukan validasi pada ketidakseimbangan data (imbalance data) sehingga diperoleh sampel training dan testing yang sesuai?
3. Bagaimana cara yang dapat digunakan untuk mengatasi ketidakseimbangan data (*imbalance data)* agar mencapai kinerja optimal?
4. Bagaimana pengaruh dari dua jenis kernel pada metode Support Vector Machine, yaitu RBF (Radial Basis Function) dan Polynomial, terhadap nilai presisi, *recall*, akurasi, dan *f-1 score*?

1.3 Batasan Masalah

Penelitian ini memasukan batasan masalah sebagai berikut:

1. Penelitian ini hanya menggunakan dataset *CICAndMal2017* yang diperoleh dari *Canadian Institute for Cybersecurity* (CIC), yaitu dataset yang terdiri dari varian SMS Malware *Jifake, Fakenotify dan Benign*.
2. Penelitian ini melakukan klasifikasi multilabel pada *SMS Malware* dengan menggunakan metode *Support Vector Machine*. Klasifikasi tersebut melibatkan varian SMS Malware Jifake, Fakenotify, dan juga kategori Benign (file normal).
3. Penelitian ini hanya menggunakan kernel RBF dan Polynomial dari metode Support Vector Machine untuk melakukan klasifikasi SMS Malware.

1.4 Tujuan

Penelitian ini memasukan tujuan sebagai berikut:

1. Melaksanakan proses klasifikasi pada *SMS Malware* dengan varian *Fakenotify, Jifake* dan *Benign* dengan salah satu jenis *machine learning* yaitu *Support Vector Machine.*
2. Melakukan proses pengambilan sampel data yang digunakan untuk keperluan *training* dan *testing*.
3. Mengimplementasikan metode SMOTE pada proses klasifikasi *SMS Malware.*
4. Menganalisis hasil kinerja dari proses klasifikasi dengan menggunakan algoritma *Support Vector Machine* dengan kernel RBF dan Polynomial.

1.5 Manfaat

Penelitian ini memasukan manfaat sebagai berikut:

1. Dataset *CICAnd Mal2017* yang berasal dari *Canadian Institute for Cybersecurity (CIC)* dapat dimanfaatkan untuk melakukan klasifikasi pasa *SMS Malware*
2. *Stratified K-Fold* memungkinkan pemberian sampel data yang tepat untuk klasifikasi *SMS Malware*.
3. SMOTE digunakan untuk menghasilkan dataset baru dari *minor class* guna mengatasi ketidakseimbangan dataset (*imbalance dataset).*
4. Validasi hasil kinerja kernel RBF dan Polynomial pada algoritma *Support Vector Machine* sebagai metode klasifikasi yang dilakukan.

1.6 Metodelogi Penelitian

Penelitian ini memasukan metodelogi penelitian meliputi yaitu mencari dan mengumpulkan referensi yang terdapat pada buku, jurnal, tesis serta sumber-sumber yang terpercaya mengenai "*Classification, SMS Malware, Support Vector Machine* dan referensi lainnya".

1. Metode Konsultasi ahli

   Metode ini dilakukan melalui proses wawancara atau konsultasi dengan para pihak yang memiliki keahlian dan pengetahuan terkait dengan permasalahan yang sedang ditangani.

2. Metode Testing Data

   Metode ini melakukan uji coba dataset yang sudah didapatkan dengan *machine learning* untuk mendapatkan hasil yang diinginkan.

3. Metode Analisa

   Pada metode ini malakukan analisa terhadap hasil pengujian yang telah diperoleh dengan cara mengamati dan mengevaluasi faktor-faktor apa yang dapat berdampak pada kinerja sistem yang dirancang.

4. Metode Kesimpulan dan Saran

   Metode ini kesimpulan dari keseluruhan kegiatan penelitian ini dan memberikan rekomendasi pada penelitian yang akan datang.

1.7 Sistematika Penulisan

Sistematikan dalam penulisan yang diajukan pada penelitian ini sebagai berikut:

BAB I PENDAHULUAN

Bab I berfokus pada latar belakang, tujuan, manfaat, perumusan masalah, batasan masalah, metodologi penelitian, dan sistematika penulisan yang digunakan dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab II memiliki isi bacaan literature untuk mendukung dan menjadi referensi penelitian yang berisi teori Malware dan metode *Support Vector Machine.*

BAB III METODOLOGI PENELITIAN

Bab III akan menjelaskan secara rinci proses-proses dalam melakukan penelitian, termasuk di dalamnya kerangka kerja, proses pembentukan sistem, langkah-langkah kerja, dan metodologi penelitian yang akan digunakan.

BAB IV HASIL DAN ANALISA

Bab IV akan Menyampaikan hasil yang diperoleh dari penelitian ini serta melakukan analisis terhadap penerapan Sistem Klasifikasi *SMS Malware* menggunakan metode Support Vector Machine.

BAB V KESIMPULAN DAN SARAN

Bab V akan dijabarkan sintesis hasil dari bab-bab sebelumnya dan disajikan saran untuk pengembangan penelitian di masa yang akan datang.

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] A. Qamar, A. Karim, and V. Chang, "Mobile malware attacks: Review, taxonomy & future directions," *Futur. Gener. Comput. Syst.*, vol. 97, pp. 887–909, 2019, doi: 10.1016/j.future.2019.03.007.

[4] 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.

[5] C. Kotkar and P. Game, "Prevention mechanism for prohibiting SMS malware attack on android smartphone," *12th IEEE Int. Conf. Electron. Energy, Environ. Commun. Comput. Control (E3-C3), INDICON 2015*, 2016, doi: 10.1109/INDICON.2015.7443579.

[6] H. Han, S. Lim, K. Suh, S. Park, S. J. Cho, and M. Park, "Enhanced android malware detection: An SVM-based machine learning approach," *Proc. - 2020 IEEE Int. Conf. Big Data Smart Comput. BigComp 2020*, no. December 2016, pp. 75–81, 2020, doi: 10.1109/BigComp48618.2020.00-96.

[7] J. Jung, K. Lim, B. Kim, S. J. Cho, S. Han, and K. Suh, "Detecting malicious android apps using the popularity and relations of APIs," *Proc. - IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. AIKE 2019*, no. 2018, pp.

309–312, 2019, doi: 10.1109/AIKE.2019.00062.

[8] X. Zhao, J. Fang, and X. Wang, “Android malware detection based on permissions,” *IET Conf. Publ.*, vol. 2014, no. 650 CP, 2014, doi: 10.1049/cp.2014.0605.

[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. 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.

[11] 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.

[12] A. Rahali, A. H. Lashkari, G. Kaur, L. Taheri, F. Gagnon, and F. Massicotte, “DIDroid: Android malware classification and characterization using deep image learning,” *ACM Int. Conf. Proceeding Ser.*, pp. 70–82, 2020, doi: 10.1145/3442520.3442522.

[13] W. Niu, R. Cao, X. Zhang, K. Ding, K. Zhang, and T. Li, “Opcode-level function call graph based android malware classification using deep learning,” *Sensors (Switzerland)*, vol. 20, no. 13, pp. 1–23, 2020, doi: 10.3390/s20133645.

[14] V. Sihag, A. Mitharwal, M. Vardhan, and P. Singh, “Opcode n-gram based malware classification in android,” *Proc. World Conf. Smart Trends Syst. Secur. Sustain. WS4 2020*, pp. 645–650, 2020, doi: 10.1109/WorldS450073.2020.9210386.

[15] H. Han, W. Y. Wang, and B. H. Mao, “Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning,” *Lect. Notes Comput.*

Sci., vol. 3644, no. PART I, pp. 878–887, 2005, doi: 10.1007/11538059_91.

[16] N. A. Diamantidis, D. Karlis, and E. A. Giakoumakis, "Unsupervised stratification of cross-validation for accuracy estimation," *Artif. Intell.*, vol. 116, no. 1–2, pp. 1–16, 2000, doi: 10.1016/S0004-3702(99)00094-6.

[17] X. Zeng and T. R. Martinez, "Distribution-balanced stratified cross-validation for accuracy estimation," *J. Exp. Theor. Artif. Intell.*, vol. 12, no. 1, pp. 1–12, 2000, doi: 10.1080/095281300146272.

[18] S. Kilgallon, L. De La Rosa, and J. Cavazos, "Improving the effectiveness and efficiency of dynamic malware analysis with machine learning," *Proc. - 2017 Resil. Week, RWS 2017*, pp. 30–36, 2017, doi: 10.1109/RWEEK.2017.8088644.

[19] B. Sanjaa and E. Chuluun, "Malware detection using linear SVM," *8th Int. Forum Strateg. Technol. 2013, IFOST 2013 - Proc.*, vol. 2, pp. 136–138, 2013, doi: 10.1109/IFOST.2013.6616872.