

***CLUSTERING ANDROID MALWARE BERDASARKAN  
FREKUENSI SYSTEM CALL MENGGUNAKAN K-MEANS***

**TUGAS AKHIR**

**Diajukan Untuk Melengkapi Salah Satu Syarat  
Memperoleh Gelar Sarjana Komputer**



**OLEH:**

**MUHAMMAD ZUFAR BADRUS  
09011381722130**

**JURUSAN SISTEM KOMPUTER  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS SRIWIJAYA  
2022**

## **LEMBAR PENGESAHAN**

### **CLUSTERING ANDROID MALWARE BERDASARKAN FREKUENSI SYSTEM CALL MENGGUNAKAN K-MEANS**

#### **TUGAS AKHIR**

**Diajukan untuk melengkapi salah satu syarat**

**Memperoleh Gelar Sarjana Komputer**

**Oleh :**

**MUHAMMAD ZUFAR BADRUS**

**09011381722130**

**Palembang, 29 Juli 2022**

**Mengetahui,**

**Ketua Jurusan Sistem Komputer**

**Pembimbing Tugas Akhir**



**A - AH**

**Ahmad Heryanto,S.Kom.M.T**  
**NIP. 198701222015041002**

# HALAMAN PERSETUJUAN

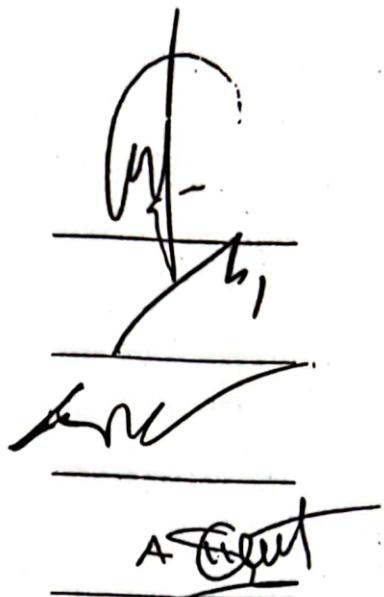
Telah diuji dan lulus pada

Hari : Jumat

Tanggal : 29 Juli 2022

Tim Penguji:

1. Ketua Sidang : Ahmad Zarkasi, M.T
2. Sekretaris Sidang : Adi Hermansyah, M.T
3. Penguji Sidang : Dr.Ir.H.Sukemi, M.T
4. Pembimbing : Ahmad Heryanto, M.T



Mengetahui

Ketua Jurusan Sistem Komputer



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

NIP 19661203200641001

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Zufar Badrus  
NIM : 09011381722130  
Program Studi : Sistem Komputer  
Judul : *Clustering Android Malware Berdasarkan frekuensi system call menggunakan K-means*

**Hasil pengecekan Software iThentivate/Turnitin : 10%**

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



Palembang 29 Juli 2022



Muhammad Zufar Badrus  
09011381722130

## **MOTTO DAN PERSEMPAHAN**

### **MOTTO:**

**Saya bisa menerima kegagalan, tapi saya tidak bisa menerima segala hal yang tak pernah diusahakan." - Michael Jordan**

### **KUPERSEMBAHKAN UNTUK:**

- ORANG TUA SAYA YANG SAYANGI DAN CINTAI YANG SELALU MENDUKUNG DAN JUGA MEMBERIKAN SEMANGAT KEPADA SAYA
- TEMAN-TEMAN SEPERJUANGAN SISTEM KOMPUTER UNIVERSITAS SRIWIJAYA ANGKATAN 2017 YANG TIDAK AKAN SAYA LUPAKAN

## 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 "*Clustering Android Malware Berdasarkan Frekuensi System Call Menggunakan K-Means*".

Dalam laporan ini penulis menjelaskan mengenai penerapan metode *System Call* dan penerapan algoritma *K-Means* untuk klastering *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 *System Call* terhadap *Malware*.

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 serta kakak dan seluruh keluarga 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 Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Ahmad Heryanto, S.Kom., M.T selaku Pembimbing Penulis 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. Teman-Teman Sistem Komputer 2017 yang sudah memberikan semangat.

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-kekangan 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, 22 Juli 2022

Penulis



***CLUSTERING ANDROID MALWARE BERDASARKAN  
FREKUENSI SYSTEM CALL MENGGUNAKAN K-MEANS***

**Muhammad Zufar Badrus (09011381722130)**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer,  
Universitas Sriwijaya

Email : baguszufar19@gmail.com

**ABSTRAK**

Pada bidang keamanan, malware yang khusus menyerang ponsel pintar berkembang semakin cepat dan canggih. *Malware* menjadi semakin kuat dalam melakukan tindakan criminal, seperti mencuri dan menghancurkan data dan informasi penting yang tersimpan di ponsel, sehingga menuntut dibuatnya sistem *antimalware* yang dapat melakukan pencegahan dan juga pendekripsi ketika terjadi serangan malware pada ponsel pintar. Dari hasil ini Berdasarkan dari penelitian menunjukkan hasil akurasi dataset tanpa seleksi fitur sebesar 50% dan dataset dengan metode seleksi fitur Sebesar 60%. Dapat disimpulkan Seleksi fitur dapat mengurangi jumlah fitur yang akan digunakan untuk keperluan klasifikasi, tetapi tidak meningkatkan hasil akurasi secara signifikan.

*Keyword : Malware, K-Means, System Call, Cluster*

**Ketua Jurusan Sistem Komputer**



**Pembimbing Tugas Akhir**

Ahmad Heryanto, S.Kom., M.T  
NIP. 198701222015041002

***CLUSTERING ANDROID MALWARE BERDASARKAN  
FREKUENSI SYSTEM CALL MENGGUNAKAN K-MEANS***

**Muhammad Zufar Badruz (09011381722130)**

Jurusan Sistem Komputer, Fakultas Ilmu Komputer,  
Universitas Sriwijaya

Email : baguszufar19@gmail.com

**ABSTRACT**

In the security sector, malware that specifically attacks smartphones is growing faster and more sophisticated. Malware is becoming more and more powerful in carrying out criminal acts, such as stealing and destroying important data and information stored on mobile phones, thus demanding the creation of an anti-malware system that can prevent and detect when carrying out malware attacks on smart phones. From these results. Based on the research, it shows that the accuracy of the dataset without feature selection is 50% and the dataset with the feature selection method is 60%. Can Lock Feature selection can reduce the number of features that will be used for classification purposes, but does not significantly increase the accuracy of the results.

*Keyword : Malware, K-Means, System Call, Cluster*

Ketua Jurusan Sistem Komputer



Dr.Ir.H.Sukemi,M.T.  
NIP. 19661203200641001

Pembimbing Tugas Akhir



AHMAD HERYANTO, S.KOM., M.T.  
NIP. 198701222015041002

## DAFTAR ISI

LEMBAR PENGESAHAN .....	ii
HALAMAN PERSETUJUAN .....	iii
HALAMAN PERNYATAAN .....	iv
MOTTO DAN PERSEMBERAHAAN .....	v
KATA PENGANTAR.....	vi
ABSTRAK.....	viii
ABSTRACT.....	ix
DAFTAR ISI .....	x
DAFTAR GAMBAR.....	xii
DAFTAR TABEL .....	xiii
DAFTAR LAMPIRAN.....	xiv
<b>BAB I</b> .....	1
1.1 Latar Belakang.....	1
1.2. Perumusan Masalah.....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan.....	3
1.5. Manfaat .....	3
1.6. Metodologi Penelitian.....	4
1.7. Sistematika Penulisan .....	5
<b>BAB II</b> .....	6
2.1. Pendahuluan .....	6
2.2. Android .....	8
2.3. Malware .....	9
2.4. System Call .....	10
2.5. Machine Learning .....	12
2.6. K-Means Clustering .....	13
2.7. Clustering Algorithms .....	16
2.8. Feature Extraction .....	19
<b>BAB III</b> .....	22
3.1. Pendahuluan .....	22
3.2. Kerangka Kerja.....	22
3.3. Raw Data .....	22

3.4 Preprocessing .....	27
3.5 Clustering K-Means .....	29
3.6 Pengujian .....	30
<b>BAB IV .....</b>	<b>33</b>
4.1. Pendahuluan .....	33
4.2. Ekstraksi Aplikasi .....	33
4.3 Praproses Data .....	37
4.4. Processing Data .....	37
4.5. Pengujian Seleksi Fitur dan Tanpa Seleksi Fitur .....	37
4.6 Hasil Pengujian K-Means Clustering .....	39
4.7 Sum Square Error .....	40
4.8. Hasil Implementasi Dengan Cross Validation .....	41
4.9. Hasil Implementasi Percentage Split .....	41
4.10. Analisa Dan Perbandingan .....	42
4.10.1. Pengujian Menggunakan data uji .....	42
4.10.2. Pengujian Menggunakan Cross Validation .....	43
4.10.3 Pengujian Menggunakan Percentage Split .....	45
<b>BAB V .....</b>	<b>51</b>
5.1 Kesimpulan .....	51
5.2 Saran .....	51
<b>DAFTAR PUSTAKA .....</b>	<b>49</b>
<b>LAMPIRAN .....</b>	<b>52</b>

## DAFTAR GAMBAR

Gambar 2. 1 Proses System Call .....	11
Gambar 2. 2 Flowchart K-Means Clustering .....	15
Gambar 2. 3 Proses K-Means Clustering .....	16
Gambar 3. 1 Flowchart Metode Penelitian .....	22
Gambar 3. 2 Proses Pengambilan System Call .....	23
Gambar 3. 3 Perintah Memulai Program ADB .....	24
Gambar 3. 4 PID Aplikasi yang diamati .....	25
Gambar 3. 5 System Call yang dipanggil .....	26
Gambar 3. 6 Alur Kegiatan Penelitian .....	28
Gambar 3. 7 Flowchart Proses Clustering K-Means .....	29
Gambar 4. 1 Proses Ekstraksi Aplikasi .....	34
Gambar 4. 2 Proses Virtualisasi .....	35
Gambar 4. 3 Proses Melakukan Pemanggilan Strace .....	36
Gambar 4. 4 Aplikasi Malware yang digunakan .....	37
Gambar 4. 5 Fitur-fitur setelah dilakukan seleksi fitur .....	38
Gambar 4. 6 Hasil Percobaan Seleksi dan Tanpa Seleksi Fitur .....	38
Gambar 4. 7 Grafik Sum Square Error dari hasil proses Clustering .....	40
Gambar 4. 9 Detail Proses Metode Cross Validator .....	43
Gambar 4. 10 ROC Table Curve Cross Validation Trojan .....	44
Gambar 4. 11 Confusion Matrix Cross Validation Trojan .....	45
Gambar 4. 12 Nilai Rata-rata dari Percentage Split .....	46
Gambar 4. 13 ROC Tabel Percentage Split .....	46
Gambar 4. 14 Confusion Matrix Pada Percentage Split .....	46

## DAFTAR TABEL

Tabel 2.1 Penelitian Terkait Yang dijadikan Rujukan.....	6
Tabel 2. 2 Metode dan Algoritma.....	12
Tabel 3. 1 Fitur-Fitur <i>System Call</i> .....	27
Tabel 3. 2 Hasil Pengujian Menggunakan Data Uji .....	30
Tabel 3. 3 Hasil Pengujian Menggunakan <i>Cross Validation</i> .....	31
Tabel 3. 4 Hasil Pengujian Menggunakan Percentage Split.....	31
Tabel 4. 1 Sum Square Error Berdasarkan Nilai K pada Proses Clustering.....	39
Tabel 4. 2 Scenario Perbandingan 3 Percobaan.....	41
Tabel 4. 3 Pengujian Pada Data Uji.....	42

## **DAFTAR LAMPIRAN**

**Lampiran 1. Data Mahasiswa**

**Lampiran 2. Hasil Cek Plagiat**

**Lampiran 3. Form Revisi Pembimbing**

**Lampiran 4. Form Revisi Pengaji**

**Lampiran 5. SULIET**

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pada bidang keamanan, malware yang khusus menyerang ponsel pintar berkembang semakin cepat dan canggih. *Malware* menjadi semakin kuat dalam melakukan tindakan criminal, seperti mencuri dan menghancurkan data dan informasi penting yang tersimpan di ponsel, sehingga menuntut dibuatnya sistem *antimalware* yang dapat melakukan pencegahan dan juga pendekripsi ketika terjadi serangan malware pada ponsel pintar [1]. Mayoritas dari Sistem pendekripsi *malware* yang telah ada menggunakan teknik pendekripsi *malware*. Dimana setiap masing-masing teknik dapat memakai pendekatan statik, dinamik ataupun *hybird* [2].

Pada penelitian [3] dilakukan di beberapa tahun terakhir dikenalkan penggunaan proses *data mining* dalam mendekripsi serangan *malware*. Beberapa diantaranya adalah melakukan penelitian dengan menggunakan metode *Data Mining*, yaitu menggunakan beberapa algoritma klasifikasi, dengan tujuan mendesain dan membangun suatu pendekripsi automatis jika terdapat malware bahkan sebelum *malware* tersebut dieksekusi. Penelitian ini membandingkan hasil yang didapatkan dengan *anti-malware* yang menggunakan metode *signature-based* tradisional.

Pada Penelitian [4] melakukan penelitian dengan mengembangkan suatu *framework* untuk melakukan klasifikasi terhadap *malware* berdasarkan informasi struktural yang dimiliki. Dari informasi struktural yang berisi fungsi-fungsi pada suatu *malware* didapatkan beberapa fitur yang digunakan untuk melakukan proses klasifikasi yaitu *Opcode*, *API*, *Memory*, *IO*, *Flag* dan *Register*.

Pada penelitian [5] mengusung teknik klastering untuk mendekripsi adanya serangan *malware*. Algoritma yang digunakan pun bermacam-macam, diantaranya K-Means dan EM clustering. Dataset

yang digunakan pada penelitian ini merupakan hasil ekstrak dari *opcode* suatu *malware*.

Selain dari penelitian-penelitian diatas, juga terdapat beberapa penelitian yang telah dilakukan. Diantara penelitian-penelitian yang dilakukan, banyak peneliti menggunakan beberapa dataset untuk melakukan pendekripsi terhadap *malware*, beberapa diantaranya menggunakan dataset berupa *system call* yaitu penelitian dari[6],[7].

Kelemahan dari Penelitian tersebut kurangnya informasi dalam perilaku perangkat lunak dalam memanfaatkan Alur Pengambilan *system call* dari malware yang sedang berjalan dalam sistem operasi ponsel pintar. Sehingga dari semua aktifitas aplikasi yang berjalan, sistem pendekripsi tidak dapat mengenali mana yang merupakan *malware* dan mana yang bukan merupakan *malware*.

Di penelitian ini memakai metode K-Means sebagai salah satu metode dari *clustering non hirarki*, cara ini membagi data ke *internal cluster*. Yang artinya bahan data yang mempunyai *character* yang mirip dapat dikelompokan ke dalam kelompok *cluster* yang sama juga data yang memiliki *character* yang beda dikelompokan pada *cluster* lain, [8] Keunggulan metode K-means disini merupakan metode analisa ataupun cara dilakukannya proses modelan tanpa supervise (unsupervised) dan juga memiliki metode yang menghasilkan kelompok data dengan sistem partisi.

Disini menggunakan *system call* sebagai bentuk komunikasi antara *user* dan *hardware* Keuntungan menggunakan *system call* di penelitian ini membuat manipulasi file bersama dengan perangkat. Seluruh device bisa digunakan seolah-olah itu adalah file system.

Relatif lebih enteng untuk mendapatkan *device driver* yang baru menggunakan *code hardware* khusus untuk mensupport file antarmuka. Maka dengan itu fungsi pengembangan *code* program pemakai ini bisa diakses melalui perangkat juga file. Yang bisa ditulis untuk mensupport API secara definisikan secara baik.

Berdasarkan latar belakang diatas, maka penulis melakukan penelitian tentang “*Clustering Android Malware Berdasarkan Frekensi System Call Menggunakan K-Means*”

## 1.2. Perumusan Masalah

1. Bagaimana penerapan seleksi fitur untuk meningkatkan efektifitas pada metode K-Means.
2. Bagaimana Mengklasifikasi Android *malware* dari pemanggilan *system call* yang sedang berjalan.

## 1.3. Batasan Masalah

Beberapa batasan masalah dalam perancangan sistem pada penelitian ini

1. Metode yang digunakan pada penelitian ini adalah metode K-Means *Clustering*.
2. Data yang digunakan dalam penelitian ini merupakan Android *malware* dengan tipe file apk dari situs penyedia malware.

## 1.4. Tujuan

Adapun tujuan dari penelitian adalah sebagai berikut:

1. Penerapan fitur seleksi menggunakan Metode *CFS (Correlation Based Feature Selection)* untuk mendapatkan hasil yang lebih baik dalam proses klasifikasi malware
2. Melakukan analisis terhadap hasil dari *Sum Square Error* dalam meningkatkan akurasi terhadap *Cluster* dari malware

## 1.5. Manfaat

Adapun manfaat dari penelitian adalah sebagai berikut:

1. Dapat mengelompokan *malware-malware* yang sedang berjalan didalam sistem operasi.
2. Dapat Mempelajari Proses *Sum Square Error* dalam mencari nilai k yang tepat

## 1.6. Metodologi Penelitian

Metodologi yang dihasilkan untuk penelitian disini melalui bermacam-macam tahap sebagai berikut:

### 1. Tahap Pertama (Perumusan masalah)

Tahap ini merupakan menentukan inti masalah tentang pengumpulan dataset. Dataset pada penelitian ini adalah data yang berisi frekuensi dari setiap sistem callyang dipanggil ketika malware dijalankan pada suatu sistem operasi.

### 2. Tahap kedua (Studi Pustaka/literatur)

Tahap ini melakukan pencarian referensi yang didapatkan dari jurnal ataupun buku yang ada kaitanya dengan metode penelitian yang berhubungan dengan tugas akhir.

### 3. Tahap ketiga (Perancangan)

Tahap ini mengandung program proses dilakukan penelitian berdasarkan rumus masalah juga literatur yang dipakai.

### 4. Tahap keempat (Pengujian)

Pada Tahap ini menguji hasil *Clustering* maka dapat diketahui bahwa malware yang digunakan dalam dataset memiliki beberapa jenis sesuai dengan klaster yang didapatkan.

### 5. Tahap kelima (Analisis)

Tahap disini menggambarkan hasil dari data yang didapatkan dan dianalisa berdasar proses clustering yang sudah dilakukan.

### 6. Tahap Keenam (Kesimpulan dan saran)

Tahap ini dilakukan dengan mengambil kesimpulan dari analisis dan studi pustaka serta membuat saran untuk dijadikan penelitian selanjutnya agar bisa dijadikan bahan referensi.

### **1.7. Sistematika Penulisan**

Penataan tulisan yang dikerjakan pada tugas akhir ini melewati beberapa tahap sebagai berikut:

## **BAB I PENDAHULUAN**

Bab satu ini mengandung tentang penjelasan secara sistematis yang topik nya diambil melingkui latar belakang, Rumusan dan Batasan masalah, tujuan dan sistematis penulisan.

## **BAB II TINJAUAN PUSTAKA**

Bab kedua disini menonjolkan penjelasan teori yang menaikan bahasan dari penelitian ini. Penjelasan teori disini mempunyai isi tentang malware, Frekuensi System Call, *Klastering Malware* dan metode sedang digunakan.

## **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan bagian tahap-tahapan penelitian yang mencakup pengujian juga analisis

## **BAB IV TUJUAN**

Pada bab ini hasil yang diperoleh dari berbagai proses pengujian kemudian dianalisis dan dijelaskan pada penelitian yang sudah dilakukan.

## **BAB V KESIMPULAN DAN SARAN**

Bab kelima disini mengangkat kesimpulan atas analisis pada penelitian yang sudah dilakukan

## DAFTAR PUSTAKA

- [1] 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.
- [2] S. Kramer and J. C. Bradfield, “A general definition of malware,” *J. Comput. Virol.*, vol. 6, no. 2, pp. 105–114, 2010, doi: 10.1007/s11416-009-0137-1.
- [3] A. Ganesan, P. Parameshwarappa, A. Peshave, Z. Chen, and T. Oates, “Extending Signature-based Intrusion Detection Systems With Bayesian Abductive Reasoning,” vol. 8, no. 5, pp. 2016–2018, 2019, [Online]. Available: <http://arxiv.org/abs/1903.12101>.
- [4] D. Kong and G. Yan, “Discriminant malware distance learning on structural information for automated malware classification,” *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, vol. Part F1288, pp. 1357–1365, 2013, doi: 10.1145/2487575.2488219.
- [5] Á. Arroyo, V. Tricio, E. Corchado, and Á. Herrero, “A comparison of clustering techniques for meteorological analysis,” *Adv. Intell. Syst. Comput.*, vol. 368, pp. 117–130, 2015, doi: 10.1007/978-3-319-19719-7\_11.
- [6] E. R. Lippincott, C. E. Weir, A. Van Valkenburg, and E. N. Bunting, “Studies of infrared absorption spectra of solids at high pressures,” *Spectrochim. Acta*, vol. 16, no. 1–2, pp. 58–73, 1960, doi: 10.1016/0371-1951(60)80071-9.
- [7] S. Hofmeyr, F. Stephanie, and S. Anil, “Intrusion detection using sequences of system calls,” *J. Comput. Secur.*, vol. 6, no. 3, pp. 151–180.
- [8] A. Bastian, H. Sujadi, and G. Febrianto, “Penerapan Algoritma K-Means Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka),” no. 1, pp. 26–32.
- [9] A. Fatima, R. Maurya, M. K. Dutta, R. Burget, and J. Masek, “Android malware detection using genetic algorithm based optimized feature selection and machine learning,” in *2019 42nd International Conference on Telecommunications and Signal Processing, TSP 2019*, Jul. 2019, pp. 220–223, doi: 10.1109/TSP.2019.8769039.

- [10] N. Idika and A.P.Mathur, “A survey of {M}alware {D}etection {T}echniques, Purdue University,” *Profsandhu.Com*2007, , [Online]. Available: [http://profsandhu.com/cs5323\\_s17/im\\_2007.pdf](http://profsandhu.com/cs5323_s17/im_2007.pdf).
- [11] S. Arshad, M. A. Shah, A. Wahid, A. Mehmood, H. Song, and H. Yu, “SAMADroid: A Novel 3-Level Hybrid Malware Detection Model for Android Operating System,” *IEEE Access*, vol. 6, 2018, doi: 10.1109/ACCESS.2018.2792941.
- [12] U. Cagliari *et al.*, “Poisoning Behavioral Malware Clustering Categories and Subject Descriptors,” pp. 27–36, 2014.
- [13] J. Saxe and D. Mentis, “Visualization of Shared System Call Sequence Relationships in Large Malware Corpora,” p. 101, 2012.
- [14] J. Soni and S. K. Peddoju, “Comparative Analysis of LSTM, One-Class SVM, and PCA to Monitor Real-Time Malware Threats Using System Call Sequences and Virtual Machine Introspection,” vol. 733, 2012, [Online]. Available: [https://link.springer.com/chapter/10.1007/978-981-33-4909-4\\_9](https://link.springer.com/chapter/10.1007/978-981-33-4909-4_9).
- [15] M. Howard, A. Pfeffer, D. M, and R. M, “Predicting signatures of future malware variants,” pp. 126–132, 2017, [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8323965/authors#authors>.
- [16] M. Kalash, N. Mohammed, M. Rochan, B. B. Neil D, and Y. Wang, “Malware Classification with Deep Convolutional Neural Networks,” 2018.<https://ieeexplore.ieee.org/abstract/document/8328749/keywords#keywords>.
- [17] M. D. Cookson and P. M. R. Stirk, “Introduction to Machine Learning with Applications in Information Security,” 2019.
- [18] I. Mahdi, “System Call Dan System Program/OS,” 2018. <https://imamansite.wordpress.com/2018/03/29/system-call-dan-system-program-os/>.
- [19] Ediyanto, Mara, M.N. & Satyahadewi, N., 2018. *Pengklasifikasikan Karakteristik Dengan Metode K-Means Cluster Analysis*. Buletin Ilmiah Mat.Stat. dan Terapannya , II(2),pp.133-36.
- [20] N. Wakhidah, “Clustering Menggunakan K-Means Algorithm ( K-Means Algorithm Clustering ),” *Fak. Teknol. Inf.*, vol. 21, no. 1, pp. 70–80, 2014.
- [21] A. Agrawal and H. Gupta, “Global K-Means (GKM) Clustering

- Algorithm: A Survey,” *Int. J. Comput. Appl.*, vol. 79, no. 2, pp. 20–24, 2013, doi: 10.5120/13713-1472.
- [22] S.Lloyd, “Least squares quantization in PCM,” pp. 129–137, [Online]. Available:<https://ieeexplore.ieee.org/abstract/document/1056489/keywords#keywords>.
- [23] J. Bilmes, “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” vol. 1198, no. 510, 2017.
- [24] M. Alazab, “Automated malware detection in mobile app stores based on robust feature generation,” *Electron.*, vol. 9, no. 3, 2020, doi: 10.3390/electronics9030435.
- [25] R. Gutierrez-osuna, “Sequential feature selection,” *Analysis*, no. 1992, pp. 1–17, 2002, [Online]. Available: [http://research.cs.tamu.edu/prism/lectures/pr/pr\\_111.pdf](http://research.cs.tamu.edu/prism/lectures/pr/pr_111.pdf).



