VISUALISASI SERANGAN MALWARE SPYWARE MENGGUNAKAN METODE *K-MEANS CLUSTERING*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer

Disusun Oleh :

Muhammad Arief Saifullah

09011381924129

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN

VISUALISASI SERANGAN MALWARE SPYWARE MENGGUNAKAN METODE *K-MEANS CLUSTERING*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu
Syarat Memperoleh Gelar Sarjana
Komputer

Oleh :

Muhammad Arief Saifullah

09011381924129

Palembang, 21/11/24

Mengetahui,

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

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Rabu

Tanggal : 13 September 2023

Tim penguji :

1. **Ketua : Sarmayanta Sembiring, S.Si., M.T**
2. **Sekretaris : Nurul Afifah, S.Kom., M.Kom**
3. **Penguji : Ahmad Heryanto, S.Kom., M.T**
4. **Pembimbing : Prof. Deris Stiawan, M.T., Ph.D**

Mengetahui 21/11/24

Ketua Jurusan Sistem Komputer

UNIVERSITAS SRIWIJAYA FAKULTAS ILMU KOMPUTER JURUSAN SISTEM KOMPUTER

Dr.Ir.Sukemi, M.T

NIP. 196612032006041001

HALAMAN PERSETUJUAN

Yang bertanda tangan di bawah ini

Nama : Muhammad Arief Saifullah

Nim : 09011381924129

Judul : VISUALIASI SERANGAN MALWARE *SPYWARE* DENGAN MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Hasil Pengecekan Sofware Tunitin : 15%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan 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 keadaan sadar dan tanpa paksaan dari siapapun.

Palembang, 18 September 2023

Muhammad Arief Saifullah
NIM. 09011381924129

HALAMAN PERSEMBAHAN

وَاِلٰى رَبِّكَ فَارْغَبْ

“ hanya kepada Tuhanmulah engkau berharap.”

(Q.S Asy-Syarh : 8)

Skripsi ini saya persembahkan untuk :

Orang tua saya tercinta Bpk. Saparudin. S.P. dan Ibu Heri Hartati Rusmeni. S.Pd. yang tidak pernah lelah dan letih mendidik saya sampai hingga saat ini, tiada hentinya juga dalam memberikan nasihat, semangat, dan arahan untuk menjadi orang sukses dan juga menjadi orang yang bermanfaat untuk banyak orang dan tak lupa juga untuk seluruh keluarga besar serta teman seperjuangan yang selalu memberikan dukungan dan semangat.

“ Jika lelah dengan sebuah proses, istirahat saja sejenak atau berjalan dengan perlahan. Jangan memutuskan untuk menyerah dan berpikir tidak ada harapan.”

-Muhammad Arief Saifullah-

VISUALISASI SERANGAN *MALWARE SPYWARE* MENGGUNAKAN METODE *K-MEANS CLUSTERING*

Muhammad Arief Saifullah (09011381924129)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : ariefsaifullah7@gmail.com

ABSTRAK

K-Means clustering adalah alat yang digunakan untuk menentukan struktur cluster dari kumpulan data yang telah diidentifikasi oleh kemiripannya yang kuat dengan cluster lain atau perbedaannya yang kuat dari cluster lain. Dalam makalah lain, dikatakan bahwa metode kerja algoritma K-Means memerlukan penggunaan centroid sebagai prototipe cluster dan hasil cluster sebelumnya sebagai outputnya. Dataset berasal dari CIC-MalMem2022 yang disediakan UNB CIC. Dataset yang diberikan CIC-MalMem2022 memiliki data yang seimbang maka tidak diperlukan lagi proses penyimbangan data. K-Means berhasil mengelompokkan 2 cluster dengan silhouette score sebesar 0,6. Hasil validasi terbaik menggunakan label K-Means dengan bantuan model Logistic Regression pada 5-Fold. Akurasi menggunakan label K-Means sebesar 99,95%, sehingga pelabelan K-Means lebih baik dibandingkan menggunakan label Malware Spyware yang hanya 99,36%.

Kata kunci : ***Malware Spyware, K-Means Clustering, Silhouette Score, Stratified K-Fold***

VISUALIZATION OF SPYWARE MALWARE ATTACKS USING K-MEANS CLUSTERING METHOD

Muhammad Arief Saifullah (09011381924129)

Computer Engineering Department, Computer Science Faculty, Sriwijaya University

Email : ariefsaifullah7@gmail.com

ABSTRACT

K-means clustering is a tool for determining the cluster structure of a data set identified by its strong similarity to other clusters or its strong differences from other clusters. Another article says that the working method of the K-Means algorithm requires using centroids as cluster prototypes and previous cluster results as output. The dataset comes from CIC-MalMem2022 provided by UNB CIC. The dataset provided by CIC-MalMem2022 has balanced data, so no data balancing process is required. K-Means managed to group 2 clusters with a silhouette score of 0.6. For best validation results, use K-Means labels using the logistic regression model on 5-Fold. The accuracy of using the K-Means label is 99.95%, so the K-Means label is better than using the malware-spyware label, which is only 99.36%.

Kata kunci : ***Malware Spyware, K-Means Clustering, Silhouette Score, Stratified K-Fold***

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur atas kehadirat Allah SWT yang telah memberikan rahmat dan karunianya sehingga penulis dapat menyelesaikan proposal Tugas Akhir yang berjudul "**Visualisasi Serangan Malware Spyware Menggunakan Metode *K-Means Clustering***"

Dalam laporan ini penulis akan memvisualisasikan dan mengklasifikasikan serangan Malware Spyware yang telah diperoleh penulis selama penelitian dan pengujian. Selain itu, penulis meyakini bahwa dengan adanya laporan tersebut maka akan sangat bermanfaat kepada khalayak umum yang ingin membaca dan tertarik untuk meneruskan penelitian tentang serangan Malware Spyware tersebut.

Sebelumnya, penulis ingin mengucapkan terima kasih kepada beberapa pihak yang telah memberikan motivasi, ide, maupun saran kepada penulis dalam penyusunan proposal Tugas Akhir ini. Untuk itu penulis ingin mengucapkan banyak terima kasih kepada:

1. Allah SWT yang telah memberikan rahmat serta karunia-Nya sehingga penulis dapat menyelesaikan Proposal Tugas Akhir ini dengan baik.
2. Orang tua saya tercinta yang telah membesarkan saya hingga saat ini dan tak henti-hentinya dalam memberikan ide, saran, serta motivasi.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Dr. Ir. Sukemi, M.T., selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Bapak Dr. Firdaus, M.Kom. selaku Dosen Pembimbing Akademik.
6. Bapak Deris Stiawan, M.T., Ph.D. selaku Dosen Pembimbing Tugas Akhir yang telah berkenan meluangkan waktunya guna membimbing, memberikan saran serta motivasi untuk penulis dalam penyusunan

Tugas Akhir ini.

7. Mbak Nurul Afifah M.Kom yang selalu memberikan arahan, masukan dan saran kepada Tim Riset *Spyware.*
8. Mbak Sari Nuzulastri selaku Admin Jurusan Sistem Komputer yang telah membantu penulis dalam hal-hal administrasi.
9. Grup Riset COMNETS.
10. Dwi Agung Laksana selaku sahabat terbaik saya yang selalu mendukung dan membantu saya dalam pembuatan penelitian ini.
11. Adi kesuma Jaya utama, Ilham Nuari, Ahmad Rifqi Akhdan yang selalu berjuang bersama selama penyusunan laporan skripsi.
12. Almamater.

Penulis sadar bahwa laporan yang disusun masih sangat jauh dari kata sempurna. Untuk itu penulis meminta kritik dan saran yang membangun sehingga agar penyusunan akan menjadi lebih baik untuk kedepannya serta menjadi daya tarik penelitian itu sendiri.

Palembang,..……

Penulis,

Muhammad Arief Saifullah

NIM. 09011381924129

DAFTAR ISI

Halaman

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1. Latar Belakang

Serangan spyware merupakan serangan dimana pencuri membobol data pribadi seseorang dan juga spyware ini dapat merusak segala sesuatu yang ada pada perangkat komputer [1]. Ketika seseorang berhasil memasukkan spyware ke dalam komputer, maka orang tersebut dapat mengakses semua data yang ada di komputer tersebut dan dapat mengakibatkan orang tersebut kehilangan data pribadinya dan juga merugikan keuangan orang tersebut.

Clustering adalah proses pengelompokan objek data ke dalam cluster yang terputus-putus sehingga data-data dikelompokkan sesuai dengan pola tersembunyi yang mungkin ditemukan dalam kumpulan data [2]. Yang termasuk dalam kluster serupa, tetapi data akan digabungkan dalam cluster yang berbeda. Banyak bidang aplikasi clustering termasuk kecerdasan buatan, biologi, manajemen hubungan pelanggan, kompresi data, penambangan data, pencarian informasi, pemrosesan gambar, pembelajaran mesin, pemasaran, kedokteran, pengenalan pola, psikologi, dan banyak lagi.

Penelitian [3], Klasifikasi hanya menggunakan algoritma C5.0 tanpa dukungan SMOTE mencapai akurasi lebih dari 90% dan kelas prediksi yang benar kurang dari 82%. Sebaliknya validasi pada contoh algoritma C5.0 dengan SMOTE k-way cross memberikan akurasi terbaik sebesar 99% dan dapat menghasilkan kelas prediksi dengan akurasi hingga 99%

Pada penelitian [4], menemukan bahwa K-Means dan FCM (Fuzzy C-Means) dapat menemukan cekungan dan jenis cluster bentuk acak lainnya ketika sebuah grup tidak dipisahkan dengan baik. FCM berperforma lebih baik pada campuran data yang noisy dalam beberapa penelitian, tetapi KM adalah pilihan yang baik untuk kumpulan data besar karena kecepatan eksekusinya. Karena akurasi dan waktu kinerjanya yang sebanding, K-Means direkomendasikan untuk analisis klaster dengan beberapa peluncur [5].

Penelitian selanjutnya [6], Algoritma KMeans++ SMOTE adalah kombinasi dari k-means++ dan algoritma SMOTE untuk menangani oversampling acak. Algoritma ini bekerja dengan memilih pusat klaster yang lebih baik dan lebih jauh, membagi pusat klaster regional, dan mengontrol laju sampel yang baru disintesis dari kelas yang berbeda. Akibatnya, hasil menunjukkan bahwa pemecah KMeans++ SMOTE berperforma lebih baik daripada pemecah lainnya (SMOTE, Borderline-SMOTE, KM-SMOTE, dan C-SMOTE) dalam menangani data yang tidak seimbang.

1.2. Rumusan Masalah

Berdasarkan latar belakang sebelumnya, penelitian ini berfokus pada distribusi cluster dan menggunakan algoritma clustering K-Means untuk menentukan titik centroid. Setelah itu, divisualisasikan antar cluster dengan algoritma koordinat paralel.

1.3. Batasan Masalah

Setelah didapatkan rumusan masalah dan latar belakang, maka berikut merupakan Batasan masalah pada penelitian ini:

1. Data yang dipakai pada penelitian ini berupa file CSV yang dimana berasal dari CIC Malmem2022.
2. Metode yang digunakan pada penelitian ini hanya menggunakan metode algoritma *K-Means Clustering*.
3. Pada penelitian ini hanya memvisualisasikan serangan *malware spyware* dan tidak membahas cara pencegahannya.

1.4. Tujuan

Adapun beberapa tujuan dari penelitian ini yaitu:

1. Memahami bagaimana pola serangan pada *malware spyware.*
2. Melakukan analisis, menentukan titik centroid dan memahami cluster pada *malware spyware.*
3. Membuat trafik pola serangan *malware spyware* kedalam bentuk grafik dan memvisualisasikannya.

1.5. Manfaat

Manfaat yang bisa didapatkan pada penelitian ini yaitut:

1. Bisa mempelajari bagaimana bentuk pola serangan *malware spyware*.
2. Bisa menjadikan referensi untuk penelitian selanjutnya tentang pencegahan serangan *malware spyware*.
3. Dapat melakukan visualisasi data dan mampu melakukan pengolahan data terhadap serangan *malware spyware*

1.6. Metodologi Penelitian

Sebelum melakukan penlitian ini dibutuhkan metode penilitian untuk melakukan penelitian ini :

1. Literatur dan studi Pustaka

   Pada tahap awal ini dilakukan dengan cara mencari beberapa referensi sumber seperti buku maupun jurnal yang dibutuhkan dan juga yang berhubungan dengan penelitian ini
2. Pemrosesan data

   Tahap ini dilakukan dengan membahas proses pengolahan data berbentuk file CSV.
3. Visualisasi

   Pada tahap ini dilakukan dengan memasukkan data ke dalam bentuk grafik dengan menggunakan parallel coordinates dan juga mencari cluster terbaik dan menetukan titik centroidnya
4. Analisa dan Kesimpulan

   Tahap ini melakukan Analisa terhadap hasil yang sudah didapatkan sebelumnya , selanjutnya tarik kesimpulan dari permasalahan dan hasil tersebut.

1.7. Sistematika Penulisan

Adapun sistematika penulisan dalam penelitian Tugas Akhir yaitu sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang penelitian dan juga menjelaskan Batasan Masalah, Rumusan Masalah, Tujuan, Manfaat, dan juga Metodologi Penelitian dan Sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi tentang Studi Pustaka yang dimana terdapat jurnal-jurnal dan penjelasan yang terkait pada penelitian ini yang bertujuan untuk sebagai referensi dasar pada penelitian.

BAB III METODOLOGI PENELITIAN

Pada Bab ini menjelaskan berbagai fase yang dilakukan peneliti selama proses penelitian

BAB IV HASIL DAN PEMBAHASAN

Bab ini merupakan pembahasan dari hasil visualisasi *cluster* dengan menggunakan *parallel coordinates* pada dataset yang digunakan dalam penelitian. Hasil dari fase sebelumnya, diuji dan diperoleh, juga didiskusikan dan dianalisis dalam bab ini, sambil memvalidasi dasar pengumpulan data yang akurat.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan yang diambil selama proses penelitian sebagai jawaban atas tujuan yang ingin dicapai, serta saran atas hasil yang diperoleh peneliti dari proses tugas akhir penelitian.

DAFTAR PUSTAKA

[1] D. Javaheri, M. Hosseinzadeh, and A. M. Rahmani, "Detection and elimination of spyware and ransomware by intercepting kernel-level system routines," *IEEE Access*, vol. 6, pp. 78321–78332, 2018, doi: 10.1109/ACCESS.2018.2884964.

[2] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[3] E. Kurniawan, F. Nhita, A. Aditsania, and D. Saepudin, "C5.0 algorithm and synthetic minority oversampling technique (SMOTE) for rainfall forecasting in bandung regency," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, vol. 4, pp. 1–5, 2019, doi: 10.1109/ICoICT.2019.8835324.

[4] Z. Cebeci and F. Yildiz, "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures," *J. Agric. Informatics*, vol. 6, no. 3, pp. 13–23, 2015, doi: 10.17700/jai.2015.6.3.196.

[5] K. Sari, S. Efendi, and S. Nasution, "Combining the Active Learning Algorithm Based on the Silhouette Coefficient with PCKmeans Algorithm," *Mecn. 2020 - Int. Conf. Mech. Electron. Comput. Ind. Technol.*, pp. 232–237, 2020, doi: 10.1109/MECnIT48290.2020.9166596.

[6] C. Li, D. Ping, S. Wei, and Z. Yan, "Improving Classification of Imbalanced Datasets Based on KM++ SMOTE Algorithm," *Proc. - 2019 2nd Int. Conf. Saf. Prod. Informatiz. IICSPI 2019*, pp. 300–306, 2019, doi: 10.1109/IICSPI48186.2019.9096022.

[7] D. Stiawan *et al.*, "Ping Flood Attack Pattern Recognition Using a K-Means Algorithm in an Internet of Things (IoT) Network," *IEEE Access*, vol. 9, pp. 116475–116484, 2021, doi: 10.1109/ACCESS.2021.3105517.

[8] S. Assitant and C. Science, "Detection of Spyware in Software," no. Icoei, pp. 1138–1142, 2019.

[9] N. A. Putri, D. Stiawan, A. Heryanto, T. W. Septian, L. Siregar, and R.

Budiarto, “Denial of service attack visualization with clustering using K-means algorithm,” *ICECOS 2017 - Proceeding 2017 Int. Conf. Electr. Eng. Comput. Sci. Sustain. Cult. Herit. Towar. Smart Environ. Better Futur.*, pp. 177–183, 2017, doi: 10.1109/ICECOS.2017.8167129.

[10] D. Javaheri, M. Hosseinzadeh, and A. M. Rahmani, “Detection and elimination of spyware and ransomware by intercepting kernel-level system routines,” *IEEE Access*, vol. 6, no. c, pp. 78321–78332, 2018, doi: 10.1109/ACCESS.2018.2884964.

[11] N. Shi, X. Liu, and Y. Guan, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” *3rd Int. Symp. Intell. Inf. Technol. Secur. Informatics, IITSI 2010*, pp. 63–67, 2010, doi: 10.1109/IITSI.2010.74.

[12] H. M. Salih and M. S. Mohammed, “Spyware Injection in Android using Fake Application,” *Proc. 2020 Int. Conf. Comput. Sci. Softw. Eng. CSASE 2020*, pp. 100–105, 2020, doi: 10.1109/CSASE48920.2020.9142101.

[13] H. Choi, H. Lee, and H. Kim, “Fast detection and visualization of network attacks on parallel coordinates,” *Comput. Secur.*, vol. 28, no. 5, pp. 276–288, 2009, doi: 10.1016/j.cose.2008.12.003.

[14] N. Ivanova, V. Gugleva, M. Dobreva, I. Pehlivanov, S. Stefanov, and V. Andonova, “Evaluation of Principal Component Analysis Variants to Assess Their Suitability for Mobile Malware Detection,” *Intech*, vol. i, no. tourism, p. 13, 2016.

[15] R. Murali, A. Ravi, and H. Agarwal, “A Malware Variant Resistant to Traditional Analysis Techniques,” *Int. Conf. Emerg. Trends Inf. Technol. Eng. ic-ETITE 2020*, pp. 1–7, 2020, doi: 10.1109/ic-ETITE47903.2020.264.

[16] Mahesh V and S. Devi K A, “Detection and Prediction of Spyware for user Applications by interdisciplinary approach,” *Int. Conf. Comput. Intell. Smart Power Syst. Sustain. Energy, CISPSSE 2020*, pp. 1–6, 2020, doi: 10.1109/CISPSSE49931.2020.9212222.

[17] F. J. Ariza-López, J. Rodríguez-Avi, and M. V. Alba-Fernández, “Complete

control of an observed confusion matrix," *Int. Geosci. Remote Sens. Symp.*, vol. 2018-July, pp. 1222–1225, 2018, doi: 10.1109/IGARSS.2018.8517540.

[18] W. Yanbo, L. Li, P. Xinfu, and F. Enpeng, "Load forecasting based on improved K-means clustering algorithm," *China Int. Conf. Electr. Distrib. CICED*, no. 201804260000067, pp. 2751–2755, 2018, doi: 10.1109/CICED.2018.8592023.

[19] H. Choi and H. Lee, "PCAV: Internet attack visualization on parallel coordinates," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3783 LNCS, no. June, pp. 454–466, 2005, doi: 10.1007/11602897_38.

[20] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical peak load clustering analysis using K-means algorithm and silhouette coefficient," *Proceeding - 2nd Int. Conf. Technol. Policy Electr. Power Energy, ICT-PEP 2020*, pp. 258–262, 2020, doi: 10.1109/ICT-PEP50916.2020.9249773.

[21] X. Zou, Y. Hu, Z. Tian, and K. Shen, "Logistic Regression Model Optimization and Case Analysis," *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2019*, pp. 135–139, 2019, doi: 10.1109/ICCSNT47585.2019.8962457.

[22] M. Dener, G. Ok, and A. Orman, "Malware Detection Using Memory Analysis Data in Big Data Environment," *Appl. Sci.*, vol. 12, no. 17, 2022, doi: 10.3390/app12178604.

[23] X. Meng, Q. Zhou, J. Hu, L. Shu, and P. Jiang, "A global support vector regression based on sorted K-fold method," *IEEE Int. Conf. Ind. Eng. Eng. Manag.*, vol. 2017-Decem, pp. 2169–2173, 2018, doi: 10.1109/IEEM.2017.8290276.

[24] S. Tricaud, K. Nance, and P. Saadé, "Visualizing network activity using parallel coordinates," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, pp. 1–8, 2011, doi: 10.1109/HICSS.2011.488.

[25] X. Pu, C. Song, and J. Huang, "Research on Optimization of Customer Value

Segmentation Based on Improved K-Means Clustering Algorithm," *Proc. 2020 IEEE 3rd Int. Conf. Inf. Syst. Comput. Aided Educ. ICISCAE 2020*, pp. 538–542, 2020, doi: 10.1109/ICISCAE51034.2020.9236867.

[26] H. Qabbaah, G. Sammour, and K. Vanhoof, "Using k-means clustering and data visualization for monetizing logistics data," *2019 2nd Int. Conf. New Trends Comput. Sci. ICTCS 2019 - Proc.*, pp. 1–6, 2019, doi: 10.1109/ICTCS.2019.8923108.

[27] "Visualisasi Serangan Malware Botnet Menggunakan Metode Clustering K-means." https://docplayer.info/220834793-Visualisasi-serangan-malware-botnet-menggunakan-metode-clustering-k-means.html.

[28] K. R. Shahapure and C. Nicholas, "Cluster quality analysis using silhouette score," *Proc. - 2020 IEEE 7th Int. Conf. Data Sci. Adv. Anal. DSAA 2020*, pp. 747–748, 2020, doi: 10.1109/DSAA49011.2020.00096.