ANALISIS KLASIFIKASI *SHELLCODE* DENGAN *MACHINE LEARNING* BERBASIS KLASIFIKASI BINER

TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Magister

OLEH:

JAKA NAUFAL SEMENDAWAI

09012682226009

PROGRAM STUDI MAGISTER ILMU KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

ANALISIS KLASIFIKASI *SHELLCODE* DENGAN *MACHINE LEARNING* BERBASIS KLASIFIKASI BINER

TESIS

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Magister

OLEH :

JAKA NAUFAL SEMENDAWAI

09012682226009

Palembang, 28 Oktober 2024

Menyetujui,

Pembimbing Utama	**Pembimbing Pendamping**
Prof. Deris Stiawan, M.T., Ph.D.	**Dr. Iwan Pahendra Anto Saputra, S.T., M.T.**
NIP. 197806172006041002	**NIP. 197403222002121002**

Mengetahui,

Ketua Prodi Magister Ilmu Komputer

Dr. Firdaus, M.Kom.

NIP. 197801212008121003

HALAMAN PERSETUJUAN

HALAMAN PERSETUJUAN

Pada hari Senin tanggal 28 Oktober 2024 telah dilaksanakan ujian sidang tesis oleh Magister Ilmu Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Jaka Naufal Semendawai

NIM : 09012682226009

Judul : Analisis Klasifikasi Shellcode Dengan Machine Learning Berbasis Klasifikasi Biner

1. Pembimbing I

   Prof. Deris Stiawan, M.T., Ph.D. .....
   NIP. 197806172006041002 1

2. Pembimbing II

   Dr. Iwan Pahendra Anto Saputra, S.T., M.T. .....
   NIP. 197403222002121002

3. Penguji I

   Dr. Abdiansah, S.Kom., M.Cs. .....
   NIP. 198410012009121005

4. Penguji II

   Dian Palupi Rini, M.Kom., Ph.D. .....

   NIP. 197802232006042002

Mengetahui,
Ketua Program Studi Magister Ilmu Komputer

Dr. Firdaus, M.Kom.
NIP. 197801212008121003

HALAMAN PERNYATAAN

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama	: Jaka Naufal Semendawai
NIM	: 09012682226009
Program Studi	: Magister Ilmu Komputer
Judul Tesis	: Analisis Klasifikasi Shellcode Dengan Machine Learning Berbasis Klasifikasi Biner

Hasil Pengecekan Software iThenticate/Turnitin : **8 %**

Menyatakan bahwa laporan tesis saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan tesis ini, maka saya bersedia menerima sanksi akademik dari universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 28 Oktober 2024

Jaka Naufal Semendawai
NIM. 09012682226009

KATA PENGANTAR

Puji dan syukur selalu kita haturkan kepada Allah SWT atas segala karunia serta nikmat yang diberikan-Nya serta shalawat serta salam selalu kita haturkan kepada junjungan Nabi Muhammad SAW, karena atas berkat, rahmat, serta karunia-Nya penulis dapat menyelesaikan Laporan Tesis yang berjudul "Analisis Klasifikasi *Shellcode* Dengan *Machine Learning* Berbasis Klasifikasi Biner". Laporan Tugas Akhir ini disusun untuk memenuhi syarat penulis untuk mendapatkan gelar Magister di Prodi Magister Ilmu Komputer Universitas Sriwijaya.

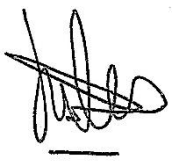
Dalam penyusunan Laporan Tugas Akhir ini, penulis menyadari banyaknya bantuan serta dukungan yang diberikan oleh banyak pihak. Ucapan terima kasih yang sebesar-besarnya penulis berikan kepada:

1. Allah SWT karena atas izin serta kemudahannya dalam menyelesaikan segala urusan selama perkuliahan.
2. Orang tua, nenek, adik, dan keluarga besar yang telah memberikan dukungan serta bantuan yang luar biasa selama ini.
3. Bapak Dr. Firdaus, M.Kom. selaku Ketua Program Studi Magister Ilmu Komputer Universitas Sriwijaya.
4. Bapak Prof. Deris Stiawan, M.T., Ph.D. selaku pembimbing utama yang telah memberikan arahan serta nasihat selama proses penyusunan Laporan Tesis ini.
5. Bapak Dr. Iwan Pahendra Anto Saputra, S.T., M.T. selaku pembimbing pendamping yang juga telah memberikan saran serta nasihat selama proses penyusunan Laporan Tesis ini.
6. Ibu Dian Palupi Rini, M.Kom., Ph.D. selaku pembimbing akademik dan dosen penguji di sidang komprehensif tesis saya yang sudah memberikan nasihat selama perkuliahan S2 ini.

7. Bapak Dr. Abdiansah, M.Cs. selaku dosen penguji di seminar proposal dan sidang komprehensif tesis saya yang sudah memberikan arahan serta masukan untuk perkembangan laporan tesis saya.
8. Seluruh dosen Program Studi Magister Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmu serta nasehat yang bermanfaat untuk saya di dunia kerja nanti.
9. Annisa Eni Salsabila, S.T. yang telah menemani serta memberikan semangat dalam proses penyusunan Laporan Tesis ini.
10. Teman-teman mahasiswa Magister Ilmu Komputer angkatan 2022 yang telah saling mendukung dalam hal apapun selama perkuliahan kurang lebih 2 tahun ini.
11. Nathan, Saeh, Ezra, Rizky, Kak Eman, dan teman-teman Kopi Eman yang sudah menjadi teman saya yang memberikan dukungan serta bantuan dalam bentuk apapun selama ini.
12. Pihak-pihak lain yang tidak bisa saya sebut satu per satu yang telah memberikan dukungan selama ini.

Penulis berharap bahwa penulisan Laporan Tesis ini dapat bermanfaat serta memberikan wawasan baru bagi pembaca walaupun penulis sadar bahwa dalam penulisan Laporan Tesis ini masih banyak kekurangan yang disebabkan oleh keterbatasan penulis. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari para pembaca. Terima kasih.

Palembang, 28 Oktober 2024



Jaka Naufal Semendawai

ABSTRAK

Internet dapat menghubungkan satu orang dengan orang lain dengan menggunakan perangkat masing-masing. Internet sendiri memiliki dampak positif dan negatif. Salah satu contoh dampak negatif dari internet adalah adanya *malware* yang dapat mengganggu atau bahkan membunuh perangkat atau penggunanya; itulah mengapa keamanan *cyber* diperlukan. Banyak metode yang dapat digunakan untuk mencegah atau mendeteksi *malware*. Salah satunya adalah dengan menggunakan teknik machine learning. Dataset pelatihan dan pengujian untuk eksperimen ini diambil dari dataset UNSW_NB15. *K-Nearest Neighbour* (KNN), *Decision Tree*, dan *Naïve Bayes* diimplementasikan untuk mengklasifikasikan apakah sebuah record pada data testing merupakan serangan *Shellcode* atau *non-Shellcode*. Pengklasifikasi KNN, *Decision Tree*, dan *Naïve Bayes* mencapai tingkat akurasi masing-masing sebesar 96,82%, 97,08%, dan 63,43%. Hasil dari penelitian ini diharapkan dapat memberikan wawasan mengenai penggunaan machine learning dalam mendeteksi atau mengklasifikasikan *malware* atau jenis serangan siber lainnya.

Kata Kunci: Klasifikasi Biner, *Cyber Security*, *Shellcode*, *Machine Learning*, *Supervised Machine Learning*, *Hyperparameter Tuning*

ABSTRACT

Internet can link one person to another using their respective devices. The internet itself has both positive and negative impacts. One example of the internet's negative impact is a malware that can disrupt or even kill a device or its users; that is why cyber security is required. Many methods can be used to prevent or detect malwares. One of the efforts is to use machine learning techniques. Training and testing dataset for the experiments is derived from the UNSW_NB15 dataset. K-Nearest Neighbour (KNN), Decision Tree, and Naïve Bayes classifiers are implemented to classify whether a record in the testing data is Shellcode or non-Shellcode attack. The KNN, Decision Tree and Naïve Bayes classifiers achieve accuracy level of 96.82%, 97.08%, and 63.43%, respectively. The results of this research are expected to provide insight into the use of machine learning in detecting or classifying malwares or other types of cyber attacks.

Key Words: Binary Classification, Cyber Security, Shellcode, Machine Learning, Supervised Machine Learning, Hyperparameter Tuning.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR RUMUS

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kesadaran akan pentingnya *cyber security* di Indonesia masih tergolong sangat rendah. Hal tersebut dibuktikan oleh data yang diterbitkan oleh *International Communication Union* (ITU) di mana tingkat *cyber security* di Indonesia ada di ranking 70. Hal tersebut menyatakan bahwa Indonesia sangat rentan untuk dikirim serangan siber dari *hacker* di negara lain. Selain itu, menurut data dari *treat exposure rate* (TER), Indonesia memiliki angka kerentanan serangan *malware* sebesar 23,54% (Ashari, 2020). Salah satu kasus *cyber security* di Indonesia adalah penyerangan yang dilakukan oleh *hacker* terhadap Pusat Data Nasional (PDN) Sementara pada bulan Juni 2024 dengan menggunakan *malware*. Hal tersebut merupakan salah satu contoh kasus yang menunjukkan keamanan data di Indonesia yang tidak baik (Saptohutomo, 2024). Hal tersebut mengakibatkan sistem pusat data nasional tersebut menjadi lumpuh (Nugroho & Wibowo, 2024).

Serangan tersebut menggunakan *malware* yang dikirimkan ke *server* dari PDN. *Malware* merupakan sebuah *software* yang dapat dioperasikan apabila seseorang berhasil masuk ke dalam ruangan *server*, lalu memasukkan *malware* tersebut menggunakan *flashdisk*. Namun saat ini, *malware* tersebut dapat dikirim melalui *internet* yang dipalsukan dalam bentuk *file* yang umum, seperti gambar, video, aplikasi, dan lain-lain (Bao et al., 2017).

Menurut Patterson et al. (2023) mengemukakan bahwa organisasi apapun sudah harus memikirkan tentang esensi dari *cyber security*. Hal tersebut dikarenakan semakin tingginya kasus serangan *cyber* yang harus dapat dilawan dengan adanya pengetahuan tentang *cyber security*, karena sudah menyangkut masalah privasi data dan ketahanan infrastruktur. Hal tersebut dapat dilihat dari riset yang dilakukan oleh Singelton et al. (2021) di mana pada tahun 2021 terdapat serangan yang menggunakan *ransomware* terhadap 10 jenis perusahaan berbeda,

dengan nilai rata-rata yaitu 17,4% dari keseluruhan jenis serangan yang terjadi pada perusahaan-perusahaan tersebut.

Salah satu *malware* yang sedang menjadi tren untuk digunakan sebagai alat serangan adalah *shellcode*. Penggunaan *shellcode* pada serangan *cyber* sudah menjadi tren di kalangan *hacker*. *Shellcode* dapat melakukan aktivitas ilegal seperti serangan DoS, pencurian data, hingga perusakan sistem secara otomatis pada komputer tujuan (Yang et al., 2022). *Shellcode* merupakan sebuah kode yang dirancang agar dapat menjalankan tugasnya secara otomatis. *Shellcode* dapat memberikan izin kepada *attacker* untuk mengeksploitasi komputer tujuan secara menyeluruh. *Shellcode* umumnya diproduksi menggunakan bahasa *assembly*. *Shellcode* dapat melakukan pengunduhan dan pengeksekusian terhadap *malware* secara otomatis ketika *hacker* berhasil masuk ke dalam sistem tujuannya. Namun, pendeteksian terhadap *shellcode* tersebut belum banyak dikembangkan.

Penelitian dari Akabane et al. (2019) yang mengembangkan metode pencegahan terhadap adanya aktivitas *shellcode* dengan hasil riset mereka yaitu *EAF Guard Driver* cukup memberikan hasil yang mengesankan bagi pencegahan *shellcode*. *Driver* tersebut dapat mencegah aktivitas *shellcode* dengan baik tanpa adanya *false alarm rate*. Dan juga hasil pengujian *benchmark* dari *EAF Guard Driver* mengalami peningkatan hingga 0,02%.

Pada penelitian yang dilakukan oleh penulis akan menggunakan *label* yang terdapat di dalam data pengujian. Sehingga, penelitian ini dilakukan dengan menggunakan *machine learning* berbasis *supervised machine learning*. Riset serupa juga dilakukan oleh Moon et al. (2022) yang menggunakan *supervised machine learning* untuk melakukan pendeteksian terhadap *malware*. Pada riset tersebut, mereka menggunakan *feature hashing* karena dapat menghemat hingga 70% memori yang digunakan, namun meningkatkan akurasi dalam pendeteksian *malware*. Sedangkan untuk penelitian ini, penulis menggunakan model lain yang termasuk ke dalam kategori *supervised machine learning* yaitu *K-nearest neighbors*, *decision tree*, dan *naive bayes* dalam melakukan klasifikasi terhadap data serangan *shellcode*.

Lalu, *binary classification* yang digunakan pada penelitian ini didasarkan oleh jumlah parameter yang digunakan untuk diteliti, yaitu keadaan normal dan keadaan setelah dimasukkan *shellcode* (Zhao & Qin, 2022). Penelitian yang dilakukan oleh Rajesh Bingu (2023) melakukan klasifikasi berbasis *binary classification* dan *multiclass classification* dengan bantuan beberapa model *machine learning*. Hasil yang didapatkan dari penelitian tersebut yaitu pada klasifikasi berbasis *binary* menghasilkan nilai akurasi dari 99,17% hingga 99,65%.

Untuk pendeteksian terhadap serangan *malware*, terdapat model yang dihasilkan dari penelitian yang dilakukan oleh Samantaray et al. (2024) yang menggunakan berbagai jenis *machine learning* seperti SVM, KNN, LR, DT, NB, dan RF. Kemudian, model tersebut menggunakan algoritma *MaxAbsScaler*. Hasil yang didapatkan dari penelitian ini yaitu nilai akurasi yang didapatkan dalam melakukan klasifikasi berbasis *multiclass classification* mengalami peningkatan dari 60% menjadi 94% dengan bantuan teknik *MaxAbsScaler*. Pada penelitian ini, penulis menggunakan metode *StandardScaler* yang digunakan pada model *K-nearest neighbors*. Hal tersebut dikarenakan metode *StandardScaler* mampu meningkatkan performa dari model KNN dalam melakukan klasifikasi terhadap data yang digunakan pada penelitian ini.

Kemudian, Gouda et al. (2024) melakukan penelitian dengan menggunakan model *decision tree* dan *k-nearest neighbors* untuk melakukan pendeteksian terhadap *malware*, dan menggunakan *data set UQ-NIDS-V2*. Di dalam *data set* tersebut terdapat serangan *shellcode*. Hasil dari penelitian tersebut menunjukkan dengan menggunakan model *decision tree* tidak mampu mendeteksi serangan *shellcode*. Hal tersebut dapat dilihat dari tingkat presisi, sensitivitas, dan *F1 Score* yaitu 0%. Namun, untuk nilai akurasi dalam pendeteksian semua jenis *malware* yaitu 98,78%. Kemudian, untuk model *K-nearest neighbors*, dalam pendeteksian *shellcode* juga tidak baik. Hal tersebut juga dapat dilihat dari tingkat presisi, sensitivitas, dan *F1 Score* yaitu 0%. Dan, untuk nilai akurasi dalam pendeteksian jenis *malware* yaitu 98,16%. Hal tersebut yang mendasari penggunaan metode *Hyperparameter Tuning* dan penskalaan data pada model *K-nearest neighbors* pada penelitian ini karena dapat meningkatkan performa dari model *K-nearest neighbors* dalam melakukan klasifikasi terhadap data yang digunakan pada penelitian ini.

Penelitian yang dilakukan untuk mendeteksi dan mencegah terjadinya serangan *malware* juga dilakukan oleh Sharma et al. (2021). Penelitian tersebut menggunakan *unsupervised machine learning* yang bertujuan untuk melakukan pendeteksian terhadap *malware*, di mana pada penelitian ini, jenis *malware* yang digunakan adalah *android ransomware*. Model yang dihasilkan dari penelitian ini disebut sebagai *RansomDroid Framework*. Model tersebut dikembangkan dari *Gaussian Mixture Model* (GMM) karena GMM dapat memberikan hasil pendeteksian yang fleksibel dan mendekati model dari *data set*. Hasil yang didapatkan dari penelitian ini yaitu tingkat akurasi sebesar 98,08% dalam 44 milisekon. Hal tersebut juga yang mendasari penggunaan *supervised machine learning* di penelitian yang penulis lakukan karena dapat meningkatkan nilai akurasi dan mempersingkat waktu dalam melakukan pendeteksian terhadap *malware*, khususnya *shellcode*.

Kemudian, terdapat penelitian yang dilakukan oleh Aljabri et al. (2024) di mana penelitian tersebut bertujuan untuk membangun model *machine learning* yang dapat mendeteksi serangan *ransomware* dengan nilai akurasi yang tinggi dan tanpa *false alarm rate*, dengan penggunaan *memory* yang seminimum mungkin. Model ini dibangun untuk dapat digunakan jika terjadi serangan yang serupa dengan *LockBit*, *Revil*, *BlackCat*, dan lain-lain. Berdasarkan pengujian yang telah dilakukan pada penelitian ini, model *XGBoost* memberikan performa terbaik, di mana menghasilkan nilai akurasi sebesar 97,58% dengan 2% *false positive rate*. Dan juga, model *XGBoost* hanya menggunakan 47 fitur dari 58 fitur yang ada. Berdasarkan penelitian tersebut, penulis ingin mengembangkan model yang berbasis *binary classification* dalam melakukan klasifikasi terhadap *malware*, khususnya *shellcode*, yang menghasilkan model dengan performa yang lebih baik dengan minim fitur yang digunakan untuk melakukan klasifikasi.

Selain itu, terdapat penelitian yang dilakukan oleh Kanemoto et al. (2019) di mana mereka menggunakan metode *shellcode emulation* yang berlandaskan kepada nilai akurasi dan performa. Mereka bertujuan untuk mendapatkan model yang dapat mengidentifikasi pemberitahuan yang penting yang dapat memberikan informasi mengenai adanya gangguan keamanan pada sistem secara otomatis. Hasil

dari penelitian ini yaitu tingkat akurasi dan performa yang didapatkan yaitu kurang lebih 60% *remote shellcode* terdeteksi.

Berdasarkan penjelasan latar belakang dan penelitian sebelumnya, maka tesis ini akan membahas tentang penggunaan model *K-nearest neighbors*, *Decision Tree*, dan *Naive Bayes* untuk melakukan klasifikasi serangan *shellcode*. Penelitian ini juga akan menerapkan metode *hyperparameter tuning* dari setiap *machine learning* yang digunakan untuk mendapatkan hasil yang lebih optimal dibandingkan dengan penelitian sebelumnya. Berdasarkan penelitian yang dilakukan oleh Muhajir et al. (2021) di mana mereka melakukan *hyperparameter tuning* pada model *machine learning* untuk melakukan klasifikasi terhadap data pendidikan. Salah satu model yang mereka gunakan adalah *K-nearest Neighbors*. Penelitian tersebut menunjukkan adanya peningkatan nilai akurasi dan *F1-Score* setelah dilakukan *hyperparameter tuning* pada model KNN, di mana nilai akurasi yang dihasilkan yaitu 82,68% dan *F1-Score* sebesar 86,58%.

Selain itu, data yang digunakan pada penelitian ini diambil dari *dataset* UNSW_NB15, di mana data di dalam *dataset* tersebut bersifat *imbalance*, sehingga menerapkan teknik *oversampling* agar dapat menyeimbangkan data dan mendapatkan hasil yang optimal juga. Berdasarkan penjabaran latar belakang di atas, maka peneliti mengangkat topik "Analisis Klasifikasi *Shellcode* Dengan *Machine Learning* Berbasis Klasifikasi Biner".

1.2. Rumusan Masalah

Berdasarkan beberapa hal yang melatarbelakangi penelitian ini, penulis mendapatkan beberapa permasalahan yang akan dirumuskan yaitu:

1. Bagaimana cara melakukan teknik *resampling* dari data *shellcode* terhadap data yang *imbalance*?
2. Bagaimana cara melakukan klasifikasi terhadap serangan *shellcode*?
3. Bagaimana mengukur performa dari metode *k-nearest neighbors, decision tree*, dan *naïve bayes*?

1.3. Batasan Masalah

Di penelitian ini, penulis membatasi beberapa parameter yang akan digunakan, di mana akan dibuat dalam poin-poin di bawah ini:

1. Teknik *resampling* yang digunakan pada penelitian ini adalah *Oversampling*.
2. Teknik *hyperparameter tuning* yang digunakan pada penelitian ini adalah *Grid Search*.
3. Untuk menentukan performa terbaik pada model *K-Nearest Neighbors* menggunakan parameter yaitu nilai K yang digunakan adalah 1 hingga 10, *metrics* yang digunakan adalah *euclidean*, dan *minkowski*, dan *weights* yang digunakan adalah *uniform* dan *distance*.
4. Pada model *naive bayes*, variasi dari pengujian terhadap data adalah penggunaan *Gaussian Naive Bayes* dan *Bernoulli Naive Bayes*. Penentuan parameter yang digunakan pada *hyperparameter tuning* yaitu nilai dari *var_smoothing* yaitu dari 10^{-9} hingga 1 pada *Gaussian Naive Bayes* dan pada *Bernoulli Naive Bayes* nilai *alpha* yaitu dari 10^{-3} hingga 10^{7}.
5. Pada model *decision tree*, variasi yang digunakan yaitu *criterion gini* dan *entropy*.

1.4. Tujuan Penelitian

Tujuan yang diharapkan untuk dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Mengetahui pengaruh dari penggunaan teknik *oversampling* terhadap data *imbalance*.
2. Mengetahui model *K-Nearest Neighbors, Decision Tree*, dan *Naïve Bayes* untuk digunakan dalam pengklasifikasian terhadap data serangan *shellcode*.
3. Mengetahui performa model *machine learning* dalam melakukan klasifikasi *shellcode*.

1.5. Manfaat Penelitian

Manfaat yang dapat diambil dari riset ini yaitu:

1. Mampu menerapkan teknik *oversampling* data *shellcode* yang *imbalance*.
2. Mampu menerapkan metode *K-Nearest Neighbors*, *Decision Tree*, dan *Naive Bayes* agar dapat digunakan untuk klasifikasi serangan *shellcode*.
3. Mampu menerapkan *confusion matrix* untuk mendapatkan data berupa akurasi dari setiap *machine learning* yang digunakan.

1.6. Sistematika Penulisan

Laporan dari penelitian ini terdiri dari 5 bab, di mana akan dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini, mayoritas berisi uraian mengenai alasan mengapa penelitian ini dilakukan yang nantinya bersifat proposal penelitian ini. Di dalam bab ini terdiri dari latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Di dalam bab ini, penulis akan membuat daftar penelitian terdahulu yang berfungsi untuk dijadikan literatur oleh penulis dalam melakukan penelitian ini. Kemudian akan terdapat studi literatur yang akan mendukung metode untuk menyelesaikan permasalahan yang muncul di penelitian ini agar metode yang digunakan bersifat kuat secara teoritis.

BAB III METODOLOGI PENELITIAN

Pada bab ini akan menguraikan kerangka kerja penelitian, data yang digunakan, kemudian metode *preprocessing* data, metode klasifikasi data, dan metode penyampaian data.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini akan diuraikan mengenai hasil klasfikasi menggunakan 3 model *machine learning*, lalu visualisasi data dari hasil pengujian, dan analisa yang dilakukan penulis terhadap data secara menyeluruh.

BAB V KESIMPULAN DAN SARAN

Pada bab ini akan menunjukkan kesimpulan serta saran dari penulis mengenai hasil analisis yang didapatkan dari bab 4.

DAFTAR PUSTAKA

Ahakonye, L. A. C., Nwakanma, C. I., Lee, J. M., & Kim, D. S. (2023). SCADA intrusion detection scheme exploiting the fusion of modified decision tree and Chi-square feature selection. *Internet of Things (Netherlands)*, *21*(December 2022), 100676. https://doi.org/10.1016/j.iot.2022.100676

Akabane, S., Miwa, T., & Okamoto, T. (2019). An EAF guard driver to prevent shellcode from removing guard pages. *Procedia Computer Science*, *159*, 2432–2439. https://doi.org/10.1016/j.procs.2019.09.418

Aljabri, M., Alhaidari, F., Albuainain, A., Alrashidi, S., Alansari, J., Alqahtani, W., & Alshaya, J. (2024). Ransomware detection based on machine learning using memory features. *Egyptian Informatics Journal*, *25*(July 2023). https://doi.org/10.1016/j.eij.2024.100445

Ashari, M. (2020). *Keamanan Informasi: Sudah Saatnya Kita Peduli*. DJKN Kemenkeu. https://www.djkn.kemenkeu.go.id/artikel/baca/13113/Keamanan-Informasi-Sudah-Saatnya-Kita-Peduli.html

Bao, T., Wang, R., Shoshitaishvili, Y., & Brumley, D. (2017). Your Exploit is Mine: Automatic Shellcode Transplant for Remote Exploits. *Proceedings - IEEE Symposium on Security and Privacy*, 824–839. https://doi.org/10.1109/SP.2017.67

Bashir, R. N., Mzoughi, O., Shahid, M. A., Alturki, N., & Saidani, O. (2024). Principal Component Analysis (PCA) and feature importance-based dimension reduction for Reference Evapotranspiration (ET0) predictions of Taif, Saudi Arabia. *Computers and Electronics in Agriculture*, *222*(May), 109036. https://doi.org/10.1016/j.compag.2024.109036

Batchu, R. K., & Seetha, H. (2021). A generalized machine learning model for DDoS attacks detection using hybrid feature selection and hyperparameter tuning. *Computer Networks*, *200*(September). https://doi.org/10.1016/j.comnet.2021.108498

Gouda, H. A., Ahmed, M. A., & Roushdy, M. I. (2024). Optimizing anomaly-based attack detection using classification machine learning. *Neural Computing and Applications*, *36*(6), 3239–3257. https://doi.org/10.1007/s00521-023-09309-y

Kanemoto, Y., Aoki, K., Iwamura, M., Miyoshi, J., Kotani, D., Takakura, H., &

Okabe, Y. (2019). Detecting successful attacks from IDS alerts based on emulation of remote shellcodes. *Proceedings - International Computer Software and Applications Conference*, *2*, 471–476. https://doi.org/10.1109/COMPSAC.2019.10251

Moon, D., Lee, J. K., & Yoon, M. K. (2022). Compact feature hashing for machine learning based malware detection. *ICT Express*, *8*(1), 124–129. https://doi.org/10.1016/j.icte.2021.08.005

Moustafa, N., Adi, E., Turnbull, B., & Hu, J. (2018). A New Threat Intelligence Scheme for Safeguarding Industry 4.0 Systems. *IEEE Access*, *6*, 32910–32924. https://doi.org/10.1109/ACCESS.2018.2844794

Moustafa, N., Creech, G., & Slay, J. (2018a). Anomaly Detection System Using Beta Mixture Models and Outlier Detection. In *Progress in Computing Analytics and Networking* (pp. 125–135). Springer. https://doi.org/https://doi.org/10.1007/978-981-10-7871-2_13

Moustafa, N., Creech, G., & Slay, J. (2018b). Flow Aggregator Module for Analysing Network Traffic. In *Progress in Computing Analytics and Networking* (pp. 19–29). Springer. https://doi.org/https://doi.org/10.1007/978-981-10-7871-2_3

Moustafa, N., Misra, G., & Slay, J. (2018). Generalized Outlier Gaussian Mixture Technique Based on Automated Association Features for Simulating and Detecting Web Application Attacks. *IEEE Transactions on Sustainable Computing*, *6*(2), 245–256. https://doi.org/10.1109/tsusc.2018.2808430

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference, MilCIS 2015 - Proceedings*. https://doi.org/10.1109/MilCIS.2015.7348942

Moustafa, N., & Slay, J. (2018). A Network Forensic Scheme Using Correntropy Variation for Attack Detection. *Advances in Digital Forensics Xiv*, 225–237.

Moustafa, N., Turnbull, B., & Choo, K.-K. R. (2018). An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things. *IEEE Internet of Things Journal*, *6*(3), 4815–4830. https://doi.org/10.1109/JIOT.2018.2871719

Muhajir, D., Akbar, M., Bagaskara, A., & Vinarti, R. (2021). Improving classification algorithm on education dataset using hyperparameter tuning. *Procedia Computer Science*, *197*, 538–544. https://doi.org/10.1016/j.procs.2021.12.171

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python*. O'Reilly Media Inc.

Nofriani, N. (2019). Comparations of Supervised Machine Learning Techniques in Predicting the Classification of the Household's Welfare Status. *Journal Pekommas*, *4*(1), 43. https://doi.org/10.30818/jpkm.2019.2040105

Nugroho, N. P., & Wibowo, E. A. (2024). *PDNS Kena Serangan Ransomware, Kominfo Pastikan Pembangunan PDN Permanen Tetap Berjalan*. Tempo.Co. https://nasional.tempo.co/read/1884499/pdns-kena-serangan-ransomware-kominfo-pastikan-pembangunan-pdn-permanen-tetap-berjalan

Pathak, A., Barman, U., & Kumar, T. S. (2024). Machine learning approach to detect android malware using feature-selection based on feature importance score. *Journal of Engineering Research (Kuwait)*, *October 2023*. https://doi.org/10.1016/j.jer.2024.04.008

Patterson, C. M., Nurse, J. R. C., & Franqueira, V. N. L. (2023). Learning from cyber security incidents: A systematic review and future research agenda. *Computers and Security*, *132*. https://doi.org/10.1016/j.cose.2023.103309

Permana, A. A., S, W., Santoso, L. W., Wibowo, G. W. N., Wardhani, A. K., Rahmaddeni, Wahidin, A. J., Yuliastuti, G. E., Elisawati, Wijayanti, R. R., & Abdurrasyid. (2023). Machine Learning. In *Machine Learning* (Vol. 45, Issue 13). https://books.google.ca/books?id=EoYBngEACAAJ&dq=mitchell+machine+learning+1997&hl=en&sa=X&ved=0ahUKEwiomdqfj8TkAhWGslkKHRCbAtoQ6AEIKjAA

Rajesh Bingu, E. al. (2023). Performance Comparison Analysis of Classification Methodologies for Effective Detection of Intrusions. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(9), 2860–2879. https://doi.org/10.17762/ijritcc.v11i9.9375

Samantaray, M., Barik, R. C., & Biswal, A. K. (2024). A comparative assessment

of machine learning algorithms in the IoT-based network intrusion detection systems. *Decision Analytics Journal*, *11*(May), 100478. https://doi.org/10.1016/j.dajour.2024.100478

Sammut, C., & Webb, G. I. (2010). Encyclopedia of Machine Learning. In *Encyclopedia of Machine Learning* (1st ed.). Springer New York. https://doi.org/10.1007/978-0-387-30164-8

Saptohutomo, A. P. (2024). *PDN Kena "Ransomware", Pemerintah Dianggap Tak Mau Belajar*. Kompas.Com. https://nasional.kompas.com/read/2024/06/27/14242581/pdn-kena-ransomware-pemerintah-dianggap-tak-mau-belajar

Sarhan, M., Layeghy, S., Moustafa, N., & Portmann, M. (2021). NetFlow Datasets for Machine Learning-Based Network Intrusion Detection Systems. *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, *371 LNICST*, 117–135. https://doi.org/10.1007/978-3-030-72802-1_9

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., & Brenning, A. (2018). *Performance evaluation and hyperparameter tuning of statistical and machine-learning models using spatial data*. https://doi.org/10.1016/j.ecolmodel.2019.06.002

Shalev, S., Shwartz, Ben, S., & David. (2014). *Understanding Machine Learning From Theory to Algorithms*. Cambridge Univesity Press.

Sharma, S., Krishna, C. R., & Kumar, R. (2021). RansomDroid: Forensic analysis and detection of Android Ransomware using unsupervised machine learning technique. *Forensic Science International: Digital Investigation*, *37*. https://doi.org/10.1016/j.fsidi.2021.301168

Singelton, C., Wikoff, A., & McMillen, D. (2021). IBM: 2021 X-Force Threat Intelligence Index. *Network Security*, 36. https://doi.org/10.1016/s1353-4858(21)00026-x

Sumaiya Thaseen, I., & Aswani Kumar, C. (2017). Intrusion detection model using fusion of chi-square feature selection and multi class SVM. *Journal of King Saud University - Computer and Information Sciences*, *29*(4), 462–472. https://doi.org/10.1016/j.jksuci.2015.12.004

Vairetti, C., Assadi, J. L., & Maldonado, S. (2024). Efficient hybrid oversampling and intelligent undersampling for imbalanced big data classification. *Expert Systems with Applications*, *246*(February 2023), 123149. https://doi.org/10.1016/j.eswa.2024.123149

Xu, Z., Shen, D., Nie, T., Kou, Y., Yin, N., & Han, X. (2021). A cluster-based oversampling algorithm combining SMOTE and k-means for imbalanced medical data. *Information Sciences*, *572*, 574–589. https://doi.org/10.1016/j.ins.2021.02.056

Yang, G., Chen, X., Zhou, Y., & Yu, C. (2022). DualSC: Automatic Generation and Summarization of Shellcode via Transformer and Dual Learning. *Proceedings - 2022 IEEE International Conference on Software Analysis, Evolution and Reengineering, SANER 2022*, 361–372. https://doi.org/10.1109/SANER53432.2022.00052

Zhang, C., Jia, D., Wang, L., Wang, W., Liu, F., & Yang, A. (2022). Comparative research on network intrusion detection methods based on machine learning. *Computers and Security*, *121*, 102861. https://doi.org/10.1016/j.cose.2022.102861

Zhao, C., & Qin, C. Z. (2022). Identifying large-area mangrove distribution based on remote sensing: A binary classification approach considering subclasses of non-mangroves. *International Journal of Applied Earth Observation and Geoinformation*, *108*(October 2021). https://doi.org/10.1016/j.jag.2022.102750