BAB I
PENDAHULUAN

1.1 Latar Belakang

Terciptanya internet pada tahun 1983 telah membawa perubahan yang sangat besar terhadap umat manusia dari era industri menjadi era digital. Perubahan era ini membawa juga tindak kejahatan di dunia nyata di dunia digital atau biasa juga disebut sebagai kejahatan siber (Cybercrime). Tindak kejahatan ini didukung dengan melunjaknya pengguna internet setiap tahunnya. Pada tahun 2011, pengguna internet mencapai sebanyak 2.14 miliar pengguna hingga sedekade kemudian pada tahun 2021 mencapai 4.88 miliar pengguna. Maka dari itu, internet tetap menjadi tempat beresiko bagi penggunanya secara digital karena memiliki potensi pengeksploitasian atas ketidaktahuan penggunanya (Kemp, 2021).

Bentuk pengeksploitasan pengguna internet biasa disebut sebagai serangan siber. Serangan siber selalu mengalami evolusi dan peningkatan kualitas dalam melakukan breach atau penjebolan dalam konteks “Attack Lifecycle”. Melihat kembali pada tahun 2022 (SonicWall Cyber Threat Report, 2023), serangan siber mengalami kenaikan sebesar 2% pada serangan berbentuk malware, ransomware mengalami menurunan sebesar 21%, ditemukan 465.501 varian malware baru melalui RTDMI atau Real-Time Deep Memory Inspection, meningkatnya Cryptojacking hingga menghasilkan $139M dalam aksinya, meningkatnya perangkat yang terhubung ke internet membuat malware pada device IoT juga turut meningkat secara signifikan sebesar 87%, percobaan intrusions juga meningkat sebesar 19% dari tahun 2021, dan serangan terenkripsi mengalami penurunan 28% dari 10.1 juta serangan di 2021 menjadi 7.3 juta serangan di 2022. Dengan angka sebesar itu, terbentuklah suatu urgensi untuk dapat mendeteksi serangan-serangan siber itu, terenkripsi atau tidak terenkripsi, secara masif dan efisien.

Tiap serangan hanya terjadi pada layer tertentu di model jaringan, seperti model jaringan OSI dan TCP/IP (Lares S, 2021). Model jaringan OSI terbagi atas 7 layer, dimulai dari paling bawah yaitu Physical layer, Data-link layer, Network layer, Transport layer, Session layer, Presentation layer, Application layer. Model jaringan populer lainnya adalah TCP/IP yang merupakan hasil gabungan beberapa

layer OSI, dimulai dari bawah yaitu Media/Network Interface layer (gabungan dari Physical layer, Data-link layer), Internet layer (di OSI itu Network Layer), Transport layer, Application layer (gabungan dari Session layer, Presentation layer, Application layer). Setiap layer ini tidak akan bisa ter-skip karena memiliki fungsinya masing-masing dan dimulai dari teratas, yaitu Application Layer yang mempunyai fungsi utama sebagai pembuat data itu (Molenaar, 2021).

Transport layer memiliki 2 protokol, yaitu TCP (Transport Control Protocol) yang memiliki ACK (acknowledgement) terhadap packets yang ia terima dan dapat melakukan recovery packets yang hilang (packets loss) dan UDP (User Datagram Protocol) yang tidak melakukan ACK, tidak bisa melakukan recovery, tetapi simple, scalable sehingga sangat cocok untuk streaming.Pada TCP, ada yang metode enkripsi yang berjalan untuk melindungi session atau data yang sedang di-transport, yaitu TLS atau Transport Layer Security (dulunya dikenal sebagai SSL atau Secure Sockets Layer dan keduanya masih dipakai hingga saat ini) (Xavier, 2021). Maka pembahasan encrypted network itu merujuk pada traffics yang terjadi di Transport Layer (*What Is TLS & How Does It Work?*, 2021).

Untuk mendeteksi adanya encrypted threat berdasarkan Sonicwall Report 2023, dapat dilakukan menggunakan machine learning ataupun deep learning. Menurut (Ferriyan et al., 2021), metode yang digunakan adalah klasifikasi untuk mengetes bagaimana kualitas dari dataset.

Tabel 1 Peforma model klasifikasi machine learning

Model Klasifikasi	**Accuracy**	**Balanced Accuracy**	**Precision**	**Recall**	**F1**
KNN	0.98	0.94	0.86	0.90	0.88
MLP	0.99	0.99	0.99	0.99	0.99
SVM	0.99	0.99	0.99	0.98	0.99
RF	0.99	0.99	0.99	0.99	0.99

Sumber: Ferriyan et al., 2021

Dapat dilihat bahwa klasifikasi machine learning RF memiliki performa yang sama dengan klasifikasi deep learning MLP. Di penelitian lain (Ferriyan, 2022), SVM

bahkan mengalahkan CNN-LSTM yang merupakan metode klasifikasi deep learning dengan sama-sama memiliki skor 99% di semua aspek tetapi SVM memiliki FPR atau False Positive Rate yang lebih rendah, yaitu SVM 0.0002 dan CNN-LSTM 0.0007. Meninjau skor-skor tadi, maka peneliti memutuskan untuk memakai klasifikasi machine learning dan juga karena resource-usage (seperti memory, cpu, dan sebagainya dalam daya komputasinya) yang besar oleh deep learning (Hwang et al., 2020).

Urgensi yang terbentuk mendorong penulis untuk melakukan penelitian yang berjudul "Aplikasi Klasifikasi Machine Learning untuk Mendeteksi Trafik Malicious di Jaringan Terenkripsi", yaitu mendeteksi keanomalian jaringan (malicious) menggunakan algoritma klasifikasi machine learning, di ruang lingkup Fakultas Teknik Universitas Hasanuddin. Setelah melihat keadaan Jaringan Fakultas Teknik Universitas Hasanuddin di Lab Cloud Computing & Ubiquitous dengan melakukan sniffing menggunakan wireshark, dapat diketahui bahwa hampir 90% traffics yang sedang berjalan adalah TLS, lebih tepatnya menggunakan HTTPS untuk kebutuhan login, checkout credit card, dan sebagainya, sama dengan yang di-report oleh (*HTTPS Encryption on the Web – Google Transparency Report*, 2023) di mana internet saat ini didominasi oleh jaringan yang terenkripsi. Maka dari itu, bisa simpulkan bahwa penelitian ini memiliki relevansi dan urgensi.

1.2 Rumusan Masalah

Untuk mendeteksi kehadiran *traffics malicious* pada jaringan terenkripsi di Kampus Fakultas Teknik Universitas Hasanuddin menggunakan pengaplikasian model klasifikasi *machine learning* diperlukan sebuah dataset serta peforma dari *machine learning* yang dipakai. Berdasarkan latar belakang tersebut, maka perlu ditanyakan:

1. Bagaimana cara menghasilkan sebuah dataset jaringan terenkripsi dari Jaringan Fakultas Teknik Universitas Hasanuddin?
2. Bagaimana performa dari model klasifikasi *machine learning* dalam mendeteksi *malicious traffics* pada jaringan Fakultas Teknik Universitas Hasanuddin?

1.3 Tujuan Penelitian

Peneliti perlu mempunyai dataset jaringan yang ditelitinya dan menggunakan model klasifikasi *machine learning* yang cocok untuk itu (Ferriyan et al., 2021). Maka dari itu tujuan penelitian yang akan dicapai adalah:

1. Untuk menciptakan sebuah dataset dengan *traffics* Jaringan Terenkripsi Fakultas Teknik Universitas Hasanuddin
2. Untuk mendeteksi keanomalian dengan menggunakan klasifikasi *machine learning* pada dataset jaringan TLS/SSL Fakultas Teknik Universitas Hasanuddin.

1.4 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini berfokus pada keamanan jaringan atau *cybersecurity*:

1. Bagi instansi, dihasilkan sebuah dataset jaringan terenkripsi yang relevan di ruang lingkup Fakultas Teknik Universitas Hasanuddin
2. Bagi pembaca umum, dapat mengetahui model klasifikasi *machine learning* yang tepat untuk mendeteksi *malicious traffics*.
3. Bagi peneliti, dapat menjadi referensi penelitian selanjutnya untuk menerapkan model klasifikasi ini pada NIDS (*Network Intrusion Detection System*) yang efektif.

1.5 Ruang Lingkup

Ruang lingkup atau batasan masalah yang dibahas oleh penelitian ini adalah sebagai berikut:

1. Batasan Ruang Lingkup:

   Untuk menciptakan lingkungan penelitian yang sama, pembaca perlu mengetahui kriteria pelaksanaan penelitian ini:

   1. Pengambilan data hanya relevan ketika menggunakan topologi jaringan yang telah ditentukan dan dilakukan di Jaringan Fakultas Teknik Universitas Hasanuddin.
   2. Hasil penelitian ini hanya relevan untuk Jaringan Fakultas Teknik Universitas Hasanuddin.

2. Batasan Objek:

   Untuk mengetahui sejauh apa penelitian ini akan mencapai tujuannya, perlu diketahui batasan objektif berikut:

   1. Penelitian ini akan memaparkan cara menghasilkan dataset jaringan terenkripsi di Jaringan Fakultas Teknik Universitas Hasanuddin
   2. Penelitian ini akan memaparkan performa model klasifikasi *machine learning*, yaitu RFC, KNN, DT, GBC, Voting, SVC, NVB.

3. Batasan Metodologi

   Untuk mengetahui bagaimana objektif itu diukur, perlu diketahui batasan metodologi pengukurannya, yaitu:

   1. Parameter model klasifikasi yang diukur adalah *cpu-profile, memory-profile, accuracy, precision, recall, F1 score, confusion matrix*

1.6 Sistematika Penulisan

Berikut adalah kerangka penulisan dari penelitian ini secara keseluruhan:

BAB I PENDAHULUAN

Pada bab ini dijelaskan mengenai latar belakang diangkatnya masalah ini menjadi penelitian, rumusan masalahnya, tujuan, manfaat, batasan masalah, metode penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini dijelaskan referensi yang merupakan landasan teori yang digunakan untuk mengimplementasi sistem, menganalisis hasil, keputusan menggunakan suatu algoritma, atau konfigurasi.

BAB III METODOLOGI PENELITIAN

Pada bab ini berisi mengenai tahapan penelitian, instrumen penelitian, pengumpulan data, perencanaan topologi, konfigurasi jaringan, penerapan algoritma, teknik pengolahan data, serta visualisasi hasil klasifikasi.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini berisi tentang sistem yang telah dibangun serta pembahasan dari hasil penelitian

BAB V PENUTUP

Pada bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.

BAB II
TINJAUAN PUSTAKA

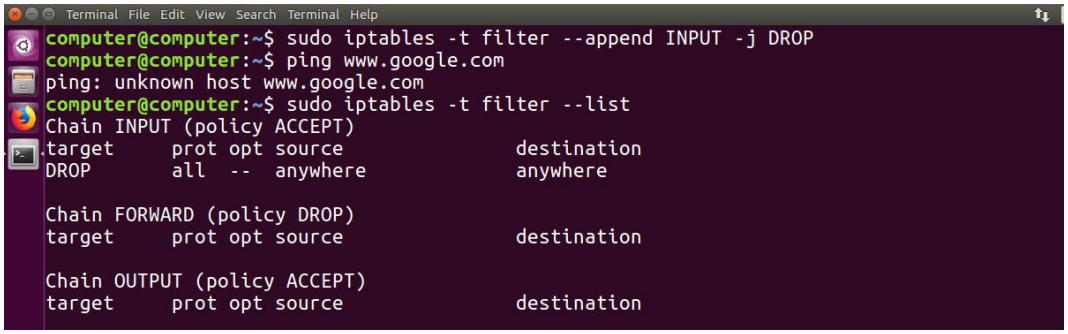
2.1 Zeek

Gambar 1 Icon dari zeek.

Zeek adalah sebuah alat analisis pasif open-source network traffics. Zeek digunakan sebagai Network Security Monitor (NSM) untuk menginvestigasi aktifitas malicious pada sebuah network traffics. Zeek dapat menjabarkan sebuah network menjadi beberapa .log files yang merekam aktifitas pada layer internet dan layer aplikasi (TCP/IP stack) (*About Zeek — Book of Zeek (v6.2.0)*, n.d.)

Zeek juga bisa sebagai framework karena dia memiliki bahasa scripting sendiri. Salah satu library dari zeek yang peneliti pakai ialah Zeek FlowMeter yang terinspirasi dari CICFlowMeter (*Zeek-Flowmeter/Zeek-Flowmeter: A Zeek Script to Generate Features Based on Timing, Volume and Metadata for Traffic Classification.*, n.d.)

2.2 IPTables



Gambar 2 Tampilan dari tabel filter Iptables (GeeksforGeeks, 2019.)

Iptables digunakan untuk mempersiapkan, maintain, dan inspect tables dari rules filter IP packet pada linux kernel (*Iptables(8) - Linux Man Page*, n.d.). Maka dari itu iptables sering digunakan menjadi *firewall* IPv4 karena ia mencocokkan *packets*

sesuai dengan rules yang sudah ditentukan oleh *user* pada iptables dan melakukan *action* terhadap *packet* tersebut. Berikut adalah beberapa istilah dalam iptables (GeeksforGeeks, 2019.):

1. *Tables* ada sebuah set dari *chains*
2. *Chain* adalah sekumpulan *rules*
3. *Rule* adalah kondisi yang akan memfilter/mencocokkan *packets*
4. *Target* adalah *action* yang dilakukan ketika rule benar; Ada ACCEPT, DROP, QUEUE
5. *Policy* adalah *default action* ketika tidak ada *rule* yang mengatur tentang *packet* tersebut pada sebuah *chains* dan bisa di ACCEPT atau DROP

Tabel 2 List tables pada Iptables

Jenis Tables	Fungsi
Filter	Default table yang digunakan untuk packet filtering, seperti chains INPUT, OUTPUT, FORWARD
NAT	Network Address Translation, seperti PREROUTING dan POSTROUTING chain
Mangle	Untuk kasus khusus packet alteration (merusak), seperti PREROUTING dan OUTPUT chain.
Raw	Untuk mem-bypass packet dari connection tracking, seperti PREROUTING dan OUTPUT chain.

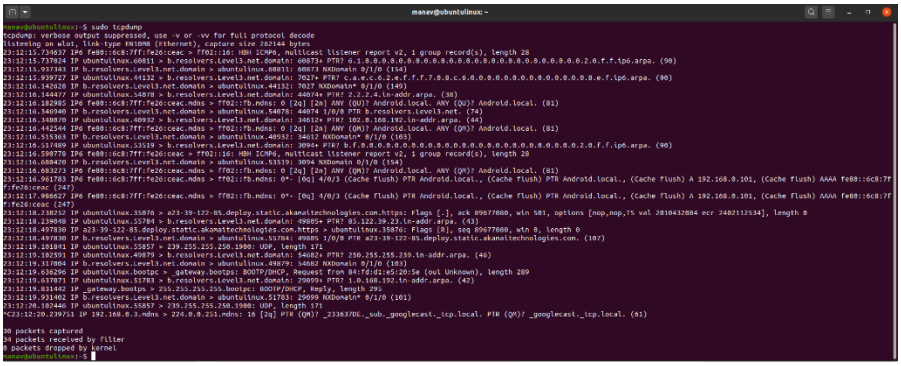
Sumber: GeeksforGeeks (2019)

Tabel 3 List chains pada Iptables

Jenis Tables	Fungsi
INPUT	Sebuah set dari rules untuk packets yang menuju local sockets
FORWARD	Untuk packets yang akan di-route melalui device ini
OUPUT	Packets yang local tadi dan akan ditransmit keluar
PREROUTING	Untuk memodifikasi packets saat diterima (arrive)
POSTROUTING	Untuk memodifikasi packets saat akan pergi (leave)
CUSTOM	Buatan user sendiri

Sumber: GeeksforGeeks (2019)

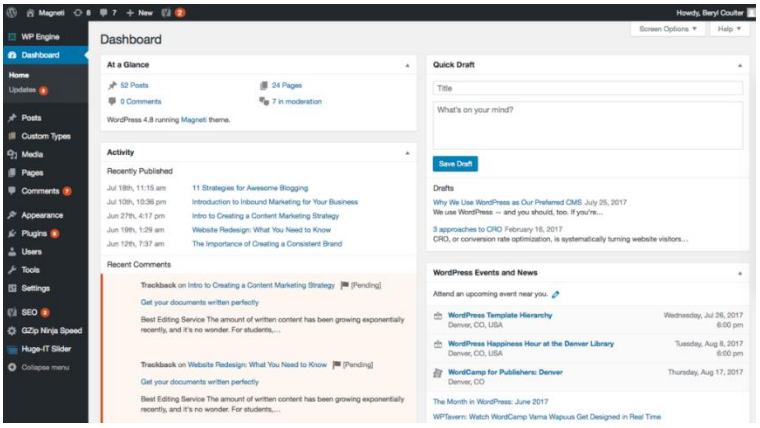
2.3 TCPdump



Gambar 3 Tampilan TCPdump di terminal linux (*Tcpdump Command in Linux with Examples - GeeksforGeeks*, n.d.)

TCPdump adalah *packet analyzer* yang melihat *data-network* pada TCP/IP dan *packets* lainnya yang sedang dikirim atau diterima. TCPdump sudah tersedia pada linux dan cocok untuk server/pc yang tidak memiliki GUI (*Graphical User Interface*) karena dia berbasis CLI (*Command Line Interface*). Proses pengambilan data ini sering disebut dengan "Dumping" dan hasilnya adalah sebuah PCAP dalam bentuk file .pcap. (*Tcpdump Command in Linux with Examples - GeeksforGeeks*, n.d.)

2.4 Web Content Management System (WCMS)

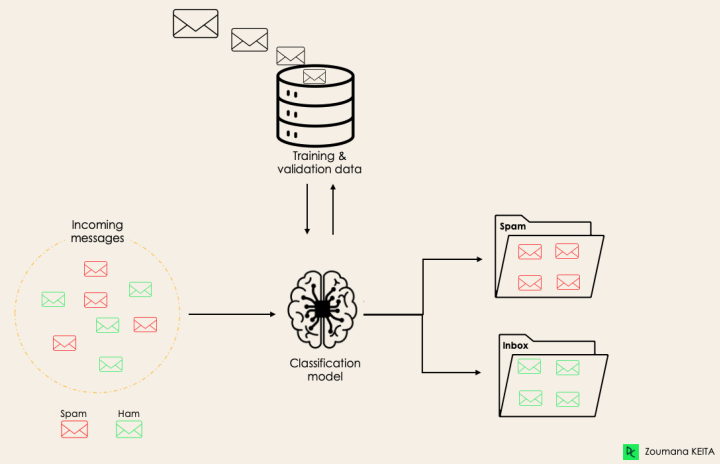


Gambar 4 Tampilan dashboard Wordpress CMS (Nicholas, 2022)

Web content management system adalah sebuah software untuk web yang dibuat untuk kepemilikan, kolaborasi, dan sistem administrasi. WCMS cocok untuk

mereka yang punya sedikit web programming atau markup languages, seperti HTML, karena dapat dengan mudah mengedit dokumen, konten, versioning, publishing, menambahkan editor baru serta mengatur level of access mereka dan banyak lagi. WCMS juga sudah integrasi dengan database, seperti CMS dari Wordpress, Drupal, dan Joomla. (Johnston, 2011)

2.5 Klasifikasi Machine Learning



Gambar 5 Contoh penggunaan model klasifikasi untuk mendeteksi email spam atau geniune (Kelta, 2022)

Menurut (Kelta, 2022) klasifikasi adalah metode supervised machine learning untuk memprediksi label dari data yang telah diinput. Model tersebut di-train menggunakan training data, kemudian dievaluasi menggunakan test data sebelum dipakai untuk memprediksi data yang belum diinput (production). Secara umum, model klasifikasi terbagi menjadi dua, yaitu

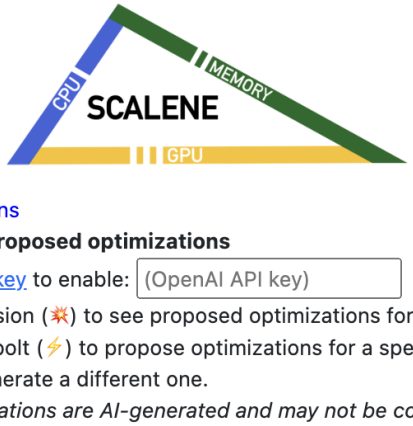
1. Eager Learners, yang lebih banyak menghabiskan waktu untuk training sebelum melakukan testing karena tujuannya untuk memahami dan membuat kesimpulan umum (berdasarkan weights). Eager Learners perlu waktu yang lama untuk men-training tapi untuk prediksi cepat. Contohnya,
    a. Logistic Regression
    b. Support Vector Machine
    c. Decision Trees

    d. Artifical Neural Network
2. Lazy Learners, yang lebih banyak menghabiskan waktu untuk memprediksi karena algoritma mereka yang “menghafal” training data dan kemudian mencari nearest neighbour atau tetangga terdekat dari data yang ingin diprediksi itu. Contohnya,
    a. K-Nearest Neighbor
    b. Case-Based reasoning

Tujuan klasifikasi terbagi atas jenis labelnya, yaitu binary classification dan multi-class classification. Binary classification adalah label yang hanya terbagi menjadi dua class, seperti binari, antara benar dan salah, sakit atau sehat, kucing atau anjing (Cristianini & Shawe-Taylor, 2000). Sedangkan, multi-class classification adalah label yang lebih dari dua class, biasanya digunakan pada Natural Language Processing, Computer Vision.

Peneliti akan mencoba semua algoritma dari *eager learners* dan *lazy learners* untuk memprediksi *traffics* jaringan kampus teknik UNHAS, yaitu *malicious* atau *benign* (*binary class*)

2.6 Scalene



Gambar 6 Logo dari library benchmarking python, Scalene
https://github.com/plasma-umass/scalene

Scalene adalah profiler CPU, GPU, dan memori berkinerja tinggi untuk Python yang melakukan sejumlah hal yang tidak dilakukan oleh profiler Python lainnya dan tidak dapat melakukannya. Ini berjalan dengan beberapa derajat kecepatan yang lebih cepat dari banyak profiler lainnya sambil memberikan informasi yang jauh lebih detail. Ini juga merupakan profiler pertama yang pernah menggabungkan optimisasi yang diusulkan oleh kecerdasan buatan. Metrik yang terekstrak dapat dilihat pada Lampiran 3