BAB I

PENDAHULUAN

1.1 Latar Belakang

Keamanan informasi merupakan suatu hal penting dalam era digital yang mengintegrasikan semua aspek ke dalam internet. Beberapa aspek yang harus dijaga dalam sebuah informasi yaitu *Confidentiality*, *Integrity*, *Avaibility*, *Authentication*, *Authorization* dan *Non Repudation* untuk memastikan bahwa informasi tersebut tidak terserang oleh pelaku kejahatan internet (Basyarahil, Astuti, dan Hidayanto 2017). Suatu informasi dapat dipastikan aman sangatlah sulit dikarenakan banyaknya penjahat internet yang berusaha untuk menyerang suatu sistem (Ramadhani 2018). Analisa dinamis dari sebuah trafik di dalam internet diperlukan untuk mengetahui ancaman serangan *malware* dengan memerhatikan perilakunya di dalam jaringan internet (Cahyanto, Wahanggara, dan Ramadana 2017). Data trafik internet yang berjumlah banyak membuat proses analisa dinamis secara manual sulit untuk dilakukan, sehingga perlu adanya sebuah algoritma *Machine Learning* yang dapat memeriksa banyak trafik sekaligus.

Proses analis *malware* memerlukan analisis yang teliti dan membutuhkan waktu yang lama untuk menemukan informasi dalam sebuah data *evidence* hasil akuisisi (Kebande dan Ray 2016). Praktik dalam proses analisa *malware* dalam dunia forensik notabenenya dilakukan oleh seorang yang awam akan teknologi yang mendalam terutama tentang digital forensik (Lakoro, Badu, dan Achir n.d.). Salah satu cara untuk mengatasi masalah tersebut adalah penerapan *machine learning* pada proses analisis data *evidence*.

Paket – paket data yang ditargetkan pada trafik *darknet* dianggap mencurigakan. Paket – paket ini sering dibuat oleh *malware* atau penyerang saat mencari target potensial berikutnya (Ban dkk., 2012). Trafik *darknet* menjadi hal yang perlu diperhatikan untuk mengetahui secara dini akan adanya ancaman serangan *malware.*

Penelitian serupa terkait penelitian ini yaitu tentang analisis ancaman *malware* dengan menggunakan dataset SURFnet yang didalamnya terdapat trafik darknet dengan menggunakan algoritma *Machine Learning* (Kumar dkk., 2019). Hasil penelitian Kumar menunjukkan bahwa dataset SURFnet masih perlu dikembangkan terkait dengan fitur dataset yang masih sedikit dan hanya memiliki dua label, hal ini diperlukan dikarenakan ancaman *malware* juga terus ikut berkembang. Yan Li dan Yu Fei juga melakukan penelitan serupa dengan menggunakan dataset CICDarknet 2020 menggunakan algoritma CNN dan LSTM yang menghasilkan akurasi yang tinggi (Li dan Lu 2021). Nilai akurasi hasil penelitian tersebut masih dapat ditingkatkan dengan menggunakan algoritma yang lain serta perlunya implementasi untuk melihat seberapa efekti CICDarknet 2020 dalam menganalisa adanya ancaman serangan malware pada trafik darknet yang telah diklasifikasi.

Penelitian dengan menggunakan Algoritma KNN pada proses klasifikasi telah banyak dilakukan diantaranya Muhamad Misbahul Azis dkk pada proses identifikasi DDOS pada jaringan SDN yang mampu melakukan proses klasifikasi mencapai nilai akurasi 0,99% (Misbahul dan Yufis 2020). Penelitian lainnya oleh

Komang dkk pada klasifikasi penggunaan protokol komunikasi pada trafik jaringan dengan nilai akurasi 0,99% (Subrata, Widyantara, dan Linawati 2016).

Latar belakang tersebut menjadi dasar dilakukannya penelitan ini yaitu untuk mengukur tingkat akurasi KNN dan melakukan seleksi fitur agar proses analisa ancaman serangan malware pada trafik *darknet* dengan menggunakan dataset CICDarknet 2020 lebih cepat.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka dapat dirumuskan masalah dalam penelitian ini adalah :

1. Bagaimana mengetahui performa akurasi penggunaan algoritma KNN pada klasifikasi trafik *darknet?*
2. Bagaimana mengetahui efektivitas kecepatan eksekusi analisa dengan menggunakan algoritma KNN pada proses klasifikasi trafik darknet*?*

1.3 Batasan Masalah

Terdapat beberapa batasan masalah yang digunakan pada penelitian ini, sebagai berikut:

1. Sumber data berasal dari data sekunder yaitu dataset CICDarknet 2020 yang dikumpulkan oleh Canadian Institute for Cybersecurity pada tanggal 19 Januari 2020 yang diunduh di laman *website* https://www.unb.ca/cic/datasets/darknet2020.html.
2. Dataset CICDarknet 2020 digunakan hanya untuk klasifikasi dan prediksi ancaman serangan *malware* berdasarkan trafik *darknet*.
3. Implementasi algoritma KNN diterapkan pada file *.pcap.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah :

1. Mengetahui performa akurasi penggunaan algoritma KNN pada klasifikasi trafik *darknet.*
2. Mengetahui efektivitas kecepatan eksekusi analisa *malware* pada trafik *darknet* dengan menggunakan algoritma KNN pada proses klasifikasi trafik darknet.

1.5 Manfaat Penelitan

Penelitan ini diharapkan dapat bermanfaat bagi seluruh pihak yang terkait, diantaraya :

1. Menjadi wawasan untuk akademisi dalam mengetahui performa penerapan algoritma *K-Nearest Neighbour* pada proses klasifikasi trafik *darknet.*
2. Membantu untuk menginvestigasi sebuah trafik darknet yang memiliki ancaman serangan malware di dalamnya dengan lebih cepat dan akurat menggunakan algoritma *KNN.*

1.6 Metodologi Penelitian

Adapun tahapan penelitian yang digunakan untuk menyelesaikan penelitian ini adalah sebagai berikut :

1. Perumusan Masalah
2. Studi Literatur
3. Pengumpulan Bahan dan Data
4. Pelabelan Data

5. Seleksi Fitur
6. Pengujian dan Pengukuran Performa KNN
7. Implementasi dan Uji Coba Sistem
8. Evaluasi dan Kesimpulan Hasil Penelitian

1.7 Struktur Penulisan Penelitan

Penulisan pada penelitian ini terbagi menjadi tiga BAB, masing - masing BAB diuraikan sebagai berikut :

Bab i pendahuluan.

Bab ini membahas latar belakang, ruang lingkup, tujuan dan manfaat penelitian, dan sistematika penelitian yang akan dilakukan untuk menjelaskan penelitian secara umum.

Bab ii landasan teori.

Bab ini menguraikan teori- teori yang berhubungan dengan proses *machine learning* yang dimana menjadi pokok utama penelitian serta *state of the art* dari penelitian sebelumnya.

Bab iii metodologi penelitian

Bab ini akan menjelaskan tentang apa saja yang akan di teliti pada penelitian ini.

Bab iv hasil dan pembahasan

Bab ini memuat analisis terhadap perancangan pada bab sebelumnya, yaitu rancangan yang sesuai dengan metodologi dan implementasi pada aplikasi yang telah dibuat, dan dilakukan pula uji coba sistem untuk mendapatkan

hasil yang sesuai dengan tujuan awal penelitian, dan terdapat kekurangan dan kelebihan dari apilkasi yang telah dibuat

Bab v kesimpulan dan saran

Bab ini merupakan bab terakhir berisi tentang kesimpulan dan saran dari hasil penelitian, serta merupakan garis besar dari metode penelitian yang telah dilakukan. Kesimpulan adalah hasil akhir dari penelitian yang dilakukan, sedangkan Saran berisi tentang rekomendasi sesuai dengan keterbatasan yang ada pada sistem usulan. Bab ini berisi kesimpulan dari hasil analisa perumusan masalah yang dirumuskan serta beberapa saran dari penulis, sehingga apa yang menjadi tujuan dari penelitian ini dapat terwujud.