BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Seiring dengan banyaknya miniaturiasi perangkat keras secara konstan, dan peningkatan efisiensi daya pada perangkat – perangkat keras tersebut di masa sekarang, telah memungkinkan untuk memasukkan kecerdasan pada perangkat keras atau elektronik yang biasa. Hal tersebut telah memunculkan istilah yang disebut *Internet of Things (IoT)*[1]. Karena istilah *IoT* ini sedang berkembang di masa sekarang, semakin banyak perangkat keras atau elektronik yang terhubung ke internet. Hal ini memungkinkan banyaknya perangkat – perangkat yang berpotensi terkena serangan *botnet*[2].

Botnet merupakan kumpulan perangkat yang terinfeksi *malware* dan dikendalikan secara *eksternal* oleh penyerang untuk melakukan serangan jaringan, seperti serangan ddos, pencurian data, atau penyebaran spam[3]. *Botnet* merupakan salah satu ancaman yang paling sering ditemukan pada sistem dan keamanaan perangkat atau jaringan *IoT* di era komputasi berbasi cloud di zaman modern[4]. Karena peningkatan yang sangat besar dalam perangkat yang saling terhubung satu sama lain, dan banyaknya sistem perangkat yang menggunakan *IoT*, jenis dan pola serangan pada *botnet* selalu berubah. Sebagai contoh, sebanyak 143.000 – 225.000 serangan *botnet IoT*

terjadi pada tahun 2018 – 2019. Untuk alasan tersebut, sangat penting untuk memahami anatomi pada *botnet*, mengklasifikasi serangan *botnet*, serta mekanisme apa yang bisa digunakan untuk menghadapi serangan berbasis *botnet* yang terjadi pada perangkat dan jaringan *IoT*[5].

Forbes menyatakan bahwa pada tahun 2019, serangan pada jaringan *IoT* terjadi sebanyak 2,9 miliar, 3 kali lebih banyak dari tahun sebelumnya. Tidak hanya itu, SonicWall Juga menyatakan bahwa serangan malware pada *IoT* meningkat 215,7% dari 10,3 juta pada tahun 2017 menjadi 32,7 juta pada tahun 2018[6]. *Machine Learning* telah digunakan dalam penelitian sebagai salah satu solusi yang berpotensi dalam menghadapi ancaman serangan *botnet* pada *IoT*, dikaranakan banyaknya data yang dihasilkan dan tersedia untuk perangkat dan jaringan *IoT*[7]. *Machine Learning* memungkinkan sistem menjadi dinamis dan fleksibel terhadap input data yang baru, dikarenakan mereka dapat "belajar" tanpa secara eksplisit diberi tahu apa yang harus dilakukan. Oleh karena itu, mereka memiliki potensi yang signifikan untuk dipakai untuk melakukan identifikasi dan klasifikasi serangan *botnet* pada jaringan *IoT*. Dalam beberapa tahun terakhir, terdapat beberapa penelitian perbandingan klasifikasi pada jaringan serangan botnet[8].

Dalam penelitian Srinivas Mukkamala dan Andrew H. Sung (2003), menggunakan machine learning untuk mendeteksi serangan botnet pada perangkat IoT menggunakan dataset UNSW-NB15. Penelitian ini mencakup eksplorasi data, pengembangan model machine learning, dan analisis

perbandingan dari berbagai model machine learning. Hasilnya menunjukkan bahwa model decision tree mendapatkan akurasi tertinggi sebesar 94%[9].

Dalam penelitian Mrutyunjaya Panda dan Mana R. Patra (2007), melakukan pengembangan model machine learning dalam mendeteksi botnet dalam IoT dengan menggunakan dataset dari N-BaIoT. Model *machine learning* ini menggunakan algoritma pemilihan fitur yang baru dan pengklasifikasi jaringan saraf yang dioptimalkan dalam mendeteksi serangan 2 botnet. Hasilnya menunjukan bahwa model FGOA-KKN mendapatkan akurasi sebesar 98.7% dalam mendeteksi serangan *botnet* dalam lingkungan IoT[10].

Dalam penelitian Dwi Widiastuti (2012), melakukan pengembangan kerangka kerja *multilayer* untuk deteksi *botnet* dengan menggunakan dataset dari CTU-13. Penelitian ini berfokus pada analisis berbasis perilaku dan fitur berbasis aliran yang efektif dalam mendeteksi lalu lintas botnet. Didapatkan performa K-NN 91.51% dan algoritma SVM mendapatkan akurasi sebesar 85%. Sehingga disimpulkan bahwa algoritma K-NN memiliki performa terbaik dalam mendeteksi aktivitas botnet dalam lalu lintas jaringan[11].

Dari penelitian – penelitian sebelumnya akurasi yang dicapai dari pengunaan metode *Support Vector Machines* untuk klasifikasi serangan botnet di atas angka 85%. Namun, untuk meningkatkan akurasi yang diperoleh pada penelitian-penelitian sebelumnya, diperlukan langkah-langkah pra-pemrosesan data yang optimal, salah satunya adalah seleksi fitur.

Seleksi fitur adalah bagian penting dalam *Machine Learning*. Seleksi fitur dapat membantu mengurangi jumlah fitur yang ada pada dataset dan memlih fitur yang paling cocok untuk melakukan klasifikasi[12]. Selain itu, seleksi fitur juga dapat mengurangi waktu komputasi algoritma ketika melakukan klasifikasi. *Mutual Information* merupakan metode untuk memilih fitur yang paling relevan dengan target *output* berdasarkan informasi bersama antara fitur dan target[13]. *Mutual Information* (MI) memiliki konsep untung menghitung seberapa banyak informasi yang terkandung dalam *term*, dan kontribusinya untuk membuat keputusan klasifikasi yang tepat pada suatu kelas. Alasan untuk membutuhkan posisi fitur yang jelas untuk memahami relevansi masing-masing fitur terhadap target output secara *eksplisit*[14].

Dalam seleksi fitur *Mutual Information* berdasarkan teori statistika, seleksi fitur menggunakan *Mutual Information* memberikan pendekatan yang kuat dan terukur untuk memilih fitur yang paling relevan terhadap target[15].

Untuk memperoleh nilai *Mutual Information* dibutuhkan beberapa nilai pendukung yaitu frekuensi *term* x pada kelas A, frekuensi *term* lain pada kelas A, frekuensi *term* x di kelas selain A, frekuensi *term* selain x di kelas selain A, dan jumlah semua *term* yang ada[16].

Seleksi fitur dilakukan untuk mereduksi fitur-fitur yang tidak relevan dalam proses klasifikasi. Seleksi fitur *Mutual Information* menggunakan teori statistika untuk menguji independensi sebuah *term* dengan kategorinya[17]. Salah satu tujuan penggunaan seleksi fitur adalah membantu fitur yang

memiliki hubungan statistika paling kuat dengan target, dengan memilih fitur yang relevan, model dapat fokus pada variabel penting yang benar-benar memengaruhi hasil[18].

Support Vector Machine adalah sebuah model yang digunakan untuk analisis regresi dan klasifikasi[19]. *Support Vector Machine* digunakan untuk masalah klasifikasi *biner* dan untuk masalah klasifikasi multikelas. *Support Vector Machine* menggunakan titik data pelatihan terdekat dari setiap data untuk membangun satu atau lebih pada *hyperplane* untuk mengklasifikasikan data berdimensi tinggi[20]. *Support Vector Machine* terdapat satu kelas mencoba menghitung *hyperplane* yang sebagian kecil dari sampel data latih terdekat dari setiap data untuk membangun *hyperplane* untuk mengklasifikasikan data berdimensi tinggi[21].

Berdasarkan penjelasan – penjelasan tersebut, maka peneliti akan melakukan penelitian untuk melakukan klasifikasi serangan *botnet* pada jaringan *Internet of Things*, dimana peneliti akan mengaplikasikan pemodelan *Machine Learning* untuk melakukan klasifikasi serangan *botnet* tersebut[22]. Pada penelitian ini, peneliti akan menggunakan metode *Machine Learning Support Verctor Machine* dengan seleksi fitur *Mutual Information*, oleh karena itu dilakukan penelitian dengan judul "Klasifikasi Botnet menggunakan *Support Vector Machines* dengan seleksi Fitur berbasis *Mutual Information* di Jaringan *Internet of Things (IoT)*"[23].

1.1 RUMUSAN MASALAH

Berdasarkan penelitian, metode *Support vector machine* telah menunjukkan hasil akurasi yang lebih baik dibanding metode *machine learning* yang lainnya dalam mengklasifikasi serangan *botnet*, yaitu sebesar 95.80%, akan tetapi, penelitian tersebut tidak menggunakan seleksi fitur dalam melakukan klasifikasi. Penelitian telah membuktikan bahwa *Mutual information* merupakan salah satu seleksi fitur yang memperhitungkan tipe data dan variabel yang diinput, dan dapat menyeleksi jumlah fitur tanpa harus mengurangi nilai akurasi pada klasifikasi, sehingga dapat mengklasifikasikan serangan *botnet* dengan baik.

Berdasarkan latar belakang yang telah dipaparkan, maka terdapat beberapa masalah yang dapat dirumuskan pada penelitian ini, yaitu:

1. Bagaimana cara mengidentifikasi fitur-fitur yang paling relevan dalam dataset botnet untuk meningkatkan akurasi serangan botnet?
2. Seberapa efektif penggunaan *Mutual Information* dalam seleksi fitur untuk meningkatkan performa algoritma *Support Vector Machines* dalam klasifikasi botnet?
3. Apa dampak dari pemilihan fitur yang optimal terhadap akurasi model SVM dalam mendeteksi berbagai jenis botnet?

1.2 BATASAN MASALAH

Adapun agar penelitian ini lebih terarah maka peneliti menetapkan batasan masalah pada penelitian ini, yaitu :

1. Algoritma yang digunakan ialah algoritma *Support Vector Machines*, dengan seleksi fitur yang digunakan ialah *Mutual Information*.
2. Dataset yang digunakan pada penelitian ialah dataset Friday-WorkingHouse-Afternoon-DDos yang di dapat dari Canadian Institute for Cybersecurity (CIC) memiliki 225.745 data serangan dan terdapat 78 atribut.
3. Penelitian ini hanya sebatas simulasi program dengan bahasa pemrograman *Python.*
4. *Output* yang dihasilkan dari penelitian ini berupa menampilkan persentase nilai *Accuracy*, *Precision*, *Recall*, dan *F1 score* dari masing – masing *Classifier*.
5. Dalam penelitian ini tidak membahas mengenai bagaimana cara pencegahan *botnet* pada jaringan *Internet of Things*.

1.3 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian ini, yaitu :

1. Mengidentifikasi fitur-fitur yang paling relevan dalam dataset botnet, dengan menggunakan seleksi fitur, sehingga dapat meningkatkan akurasi dalam mendeteksi serangan botnet.
2. Mengevaluasi sejauh mana efektrifitas penggunaan metode *Mutual Information* dalam proses seleksi fitur dapat meningkatkan performa

algoritma *Support Vetor Machines* dalam melakukan klasifikasi serangan botnet.

3. Mengukur nilai performa algoritma *Support Vector Machines* dalam mengkalsifikasi serangan *botnet*.

1.4.2 Manfaat Penelitian

Adapun manfaat dari penelitian ini :

1. Identifikasi fitur- fitur yang paling relevan dapat meningkatkan kinerja model dalam mendeteksi serangan botnet.
2. Meningkatkan akurasi sejauh mana metode *Mutual Information* dan algoritma *Support Vector Machines* dalam mengklasifikasi serangan botnet.
3. Mengetahui nilai performa dari algoritma *Support Vector Machines* dalam klasifikasi serangan *Botnet*.

1.4 SISTEMATIKA PENULISAN

Sistematika dibuat untuk memberi suatu gambaran yang jelas terhadap penulisan ilmiah agar menghindari terjadinya pembahasan diluar dari tema dan judul penelitian. Secara garis besar penulisan ilmiah terdiri dari 5 bab, dapat dilihat dari sistematika sebagai berikut :

- **BAB I : PENDAHULUAN**

  Bab ini merupakan bab pendahuluan yang membahas tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan dan manfaat

penelitian, metodologi penelitian, jadwal penelitian serta sistematika penulisan.

- **BAB II : LANDASAN TEORI**

  Bab ini memuat tentang landasan teori yang mendasari pembahasan laporan secara khusus dan berisikan definisi-definisi yang melandasi penelitian yang dilakukan dengan melakukan studi literature sebagai dasar dalam melakukan analisa dan perancangan.

- **BAB III : METODOLOGI PENELITIAN**

  Bab ini membahas secara rinci kerangka kerja penelitian, metode atau pendekatan yang digunakan, dan *tools* (alat bantu) yang digunakan menjawab masalah penelitian untuk mencapai sebuah tujuan.

- **BAB IV : ANALISIS PENGUJIAN**

  Bab ini menyajikan pembahasan hasil yang diperoleh dan model deteksi menggunakan klasifikasi yang dibangun dalam penelitian ini. Serta evaluasi dan pengujian sistem yang telah dihasilkan untuk melihat kesesuaian dengan tujuan yang diharapkan.

- **BAB V : PENUTUP**

  Bab ini berisi kesimpulan dan saran-saran daripenelitian guna mengembangkan analisa deteksi serangan yang telah diuji coba.

DAFTAR PUSTAKA

[1] Y. I. Mahendra and R. E. Putra, "Penerapan Algoritma Gradient Boosted Decision Tree (GBDT) untuk Klasifikasi Serangan DDoS," *JINACS (Journal Informatics Comput. Sci. ISSN*, vol. 06, pp. 158–166, 2024.

[2] Y. Prihantono, M. T. Yusuf, A. Kurniawan, A. Ilarizky, I. Komputer, and U. Borobudur, "ANALISA PERBANDINGAN ALGORITMA MACHINE LEARNING DECISION TREES DAN EXTRA TREES TERHADAP KINERJA INTRUSION DETECTION SYSTEM DALAM MEMPREDIKSI," vol. 10, no. 1, pp. 53–63, 2024.

[3] A. Nugraha and F. A. Rafrastara, "Botnet Detection Survey," *Semin. Nas. Teknol. Inf. Komun. Terap. 2011*, vol. 1, no. Semantik, 2011.

[4] R. Firdaus, A. I. Hadiana, and F. Kasyidi, "Model Deteksi Botnet Menggunakan Algoritma Decision Tree Dengan Untuk Mengidentifikasi Serangan Click Fraud," *J. Informatics Commun. Technol.*, vol. 4, no. 2, pp. 10–20, 2023, doi: 10.52661/j_ict.v4i2.122.

[5] G. Martha and G. Bororing, "Evaluasi Kinerja Algoritma Machine Learning Dalam Prediksi Serangan Malware," *J. Rev. Pendidik. dan Pengajaran*, vol. 7, no. 1, pp. 3060–3066, 2024.

[6] S. Susanto, M. A. S. Arifin, and H. O. L. Wijaya, "IoT Botnet Detection Using Autoencoders and Decision Trees," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 12, no. 3, pp. 329–334, 2023, doi:

10.32736/sisfokom.v12i3.1693.

[7] M. Panda and M. R. Patra, “Detecting Network Intrusions : A Clustering Approach Detecting Network Intrusions - A Clustering Approach,” no. January 2009, 2014.

[8] M. Agus Syamsul Arifin, A. Anto Tri Susilo, A. Taqwa Martadinata, and B. Santoso, “KLIK: Kajian Ilmiah Informatika dan Komputer Deteksi Aktifitas Malware pada Internet of Things menggunakan Algoritma Decision Tree dan Random Forest,” *Media Online)*, vol. 4, no. 6, pp. 3073–3079, 2024, doi: 10.30865/klik.v4i6.1903.

[9] A. Rachmat, “Survei Penerapan Model Machine Learning Dalam Bidang Keamanan Informasi,” *J. Sist. Cerdas*, vol. 2, no. 1, pp. 47–60, 2019, doi: 10.37396/jsc.v2i1.20.

[10] N. Y. Abel Filemon Haganta Kaban, Indriati, “Analisis Sentimen Aplikasi E-Goverment berdasarkan Ulasan Pengguna menggunakan Metode Maximum Entropy dan Seleksi Fitur Mutual Information,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 1452–1458, 2021.

[11] A. Selay *et al.*, “Karimah Tauhid, Volume 1 Nomor 6 (2022), e-ISSN 2963-590X,” *Karimah Tauhid*, vol. 1, no. 2963–590X, pp. 861–862, 2022.

[12] M. Zidane, “Klasifikasi Serangan Distributed Denial-of-Service (DDoS) menggunakan Metode Data Mining Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 1, pp. 172–180, 2022, [Online]. Available: http://j-ptiik.ub.ac.id

[13] “SIMPLE BOTNET IN C PROGRAMMING TUGAS BESAR Mata Kuliah Jaringan Komputer Lanjut Institut Teknologi Bandung FAISAL AKBAR (23216099) NURHUDA MAULANA (23216012) SURIANDI JAYAKRISTA (23215360) Program Studi Magister Teknik Elektro SEKOLAH TEKNIK ELEKTRO,” no. 23216099, 2017.

[14] U. Maikudi, O. A. Abisoye, S. O. Ganiyu, and S. A. Bashir, “A Literature Survey on IoT Botnet Detection Techniques,” p. 110, 2021.

[15] M. Lefoane, I. Ghafir, S. Kabir, and I. U. Awan, “Machine Learning for Botnet Detection: An Optimized Feature Selection Approach,” *ACM Int. Conf. Proceeding Ser.*, pp. 195–200, 2021, doi: 10.1145/3508072.3508102.

[16] R. F. A. Bintoro, P. H. Trisnawan, and M. Data, “Deteksi Bot Network (BOTNET) menggunakan Metode Decision Tree dari Dataset CTU,” *... Teknol. Inf. dan Ilmu ...*, vol. 7, no. 6, pp. 2921–2930, 2023, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12857

[17] S. Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, “Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 1, no. 2, p. 180, 2017, doi: 10.30645/j-sakti.v1i2.41.

[18] S. Haryati, A. Sudarsono, and E. Suryana, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.

[19] D. Irawan, E. B. Perkasa, Y. Yurindra, D. Wahyuningsih, and E. Helmud, "Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 10, no. 3, pp. 432–437, 2021, doi: 10.32736/sisfokom.v10i3.1302.

[20] H. Apriyani and K. Kurniati, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.isssue3.year2020.page133-143.

[21] A. Wibowo, "Uji Chi-Square pada Statistika dan SPSS," *J. Ilm. SINUS*, vol. 4, no. 2, p. 38, 2017.

[22] A. Hidayat, Y. Samudra, and L. P. Andriyanto, "Sosialisasi Pengenalan Pentingnya Cyber SecurityBagi Siswa Untuk Membangun Keamanan Informasi Dalam Era Digital," *J. Pengabdi. Masy.*, vol. 2, no. 5, pp. 450–457, 2023.

[23] M. K. Harto and A. Basuki, "Deteksi Serangan DDoS Pada Jaringan Berbasis SDN Dengan Klasifikasi Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, pp. 1329–1333, 2021, [Online]. Available: http://j-ptiik.ub.ac.id