BAB I
PENDAHULUAN

1.6 Latar Belakang

Internet merupakan salah satu teknologi yang berevolusi sangat pesat, sehingga sampai bisa menawarkan banyak fasilitas yang bisa digunakan secara bebas. Cakupan yang luas dan fleksibel menjadi daya tarik yang luar biasa tinggi, apalagi dengan akses yang mudah membuat hampir semua kalangan bisa memanfaatkan fasilitas internet yang tersedia. Semakin banyaknya pengguna maka semakin banyak informasi yang tersimpan. Celah inilah yang dimanfaatkan orang yang tidak bertanggung jawab untuk melakukan berbagai serangan digital atau kejahatan siber. Salah satu teknik kejahatan siber tersebut adalah DDoS. DDoS adalah ancaman terhadap keamanan jaringan yang bertujuan untuk melemahkan kinerja lalu lintas jaringan menggunakan lalu lintas palsu atau berbahaya [1]. Selain jaringan, server juga bisa menjadi bahan target untuk penyerang melakukan DDoS. Walaupun serangan DDoS sudah ada sejak lama, teknik serangan ini masih sering digunakan oleh para penyerang. Laporan dari situs Cloudflare [2], terjadi serangan pada server Minecraft di kuartal ketiga tahun 2022, Mirai botnet menjadi serangan terbesar. Ini menjadi bukti bahwa serangan DDoS masih relevan dan bisa mengikuti perkembangan teknologi dan masih banyak digunakan oleh para penindak kejahatan siber.

Dari beberapa penelitian sebelumnya, sudah banyak yang mengembangkan beberapa dataset serangan DDoS untuk dijadikan latihan dan menguji pada machine learning. The 1999 DARPA/Lincoln Laboratory IDS Evaluation Data [3], salah

satu dataset yang sudah sering digunakan para peneliti dan komunitas jaringan untuk deteksi intrusi. Pada penelitian [4] berhasil mendeteksi dan menghasilkan dataset DDoS menggunakan Enhanced Multi Class Support Vector Machine (EMCSVM). Pada penelitian [5] mencoba melakukan simulasi serangan DDoS dengan cara dua jenis yaitu serangan vulnerability(kelemahan) dan serangan flooding(membanjiri). Pada dataset CAIDA "DDoS 2007 Attack." [6] berisi jejak lalu lintas jaringan selama satu jam yang mana dilakukan secara anonim menggunakan CryptoPAn. Lalu pada penelitian [7] dengan mempelajari dan menggunakan perangkat yang berfokus pada layer aplikasi seperti RUDY dan slowloris, mereka dapat mengekstrasi parameter dari tangkapan paket yang berfungsi sebagai tanda serangan DDoS. Tetapi dikarenakan banyak kekurangan dan masalah pada dataset sebelumnya, seperti lalu lintas yang tidak lengkap, data yang dianonimkan dan skenario serangan yang sudah kuno dapat menyebabkan pembatasan pengujian dan evaluasi model deteksi yang diusulkan [1].

Oleh sebab itu dibutuhkan dataset yang komprehensif untuk menunjang penelitian yang efektif dan tetap relevan. Pada penelitian [1] telah melakukan simulasi dan riset terhadap serangan DDoS dan akhirnya menghasilkan sebuah dataset komprehensif yaitu CIC-DDoS2019. Pada dataset CIC-DDoS2019 berhasil menganalisa lalu lintas jaringan dengan menggunakan CICFlowMeter-V3. Dataset tersebut berisikan benign(normal) dan beberapa serangan DDoS umum yang paling terbaru, yang mana bisa mirip dengan data nyata yang sebenarnya(PCAPs). Di dalam dataset CIC-DDoS2019 terdapat imbalanced(tidak seimbang) data yang harus dihadapi untuk melakukan riset. Maka dari itu peneliti berinisiatif untuk

melakukan riset klasifikasi identifikasi serangan DDoS menggunakan seleksi fitur cfs-pso search pada algoritma klasifikasi J48, Naive Bayes dan Radial Basis Function Network untuk meningkatkan akurasi deteksi serangan DDoS dari penelitian sebelumnya.

1.6 Rumusan Masalah

1. Apakah seleksi fitur mempengaruhi kinerja algoritma klasifikasi?
2. Fitur-fitur apa saja yang relevan terhadap kinerja klasifikasi?
3. Adakah peningkatan kinerja klasifikasi setelah dilakukan seleksi fitur?
4. Adakah penurunan kinerja klasifikasi setelah dilakukan seleksi fitur?

1.6 Batasan Masalah

Dalam pelaksanaan penelitian ini, berbagai permasalahan yang muncul dalam konteks objek yang lebih luas akan dibatasi sesuai dengan kebutuhan dan kemampuan. Dalam hal ini batasan masalah dalam penelitian ini antara lain:

1. Dataset untuk training dan testing didapatkan dari repository University of New Brunswick : DDoS Evaluation Dataset (CIC-DDoS2019).
2. Pada penelitian ini menggunakan seleksi fitur berbasis CFs-PSO search kemudian fitur pada dataset dilakukan klasifikasi menggunakan algoritma klasifikasi. Dengan batas hasil akurasi setelah seleksi atribut.
3. Metode yang digunakan untuk klasifikasi adalah J48, Naive Bayes, dan Radial Basis Function Neural Network
4. Teknik Confusion Matrix digunakan untuk menghitung tingkat akurasi.

5. Pengolahan data dilakukan dengan menggunakan aplikasi WEKA versi 3.8.6.

1.6 Tujuan Penelitian

Dalam pembuatan penelitian ini, adapun tujuan adalah sebagai berikut :

1. Menerapkan seleksi fitur berbasi CFs-PSO Search untuk menguji pengaruh pada kinerja algoritma klasifikasi.
2. Menentukan fitur-fitur yang relevan berdasarkan hasil seleksi fitur berbasis CFs-PSO search.
3. Mengetahui peningkatan kinerja klasifikasi setelah diterapkan seleksi fitur.
4. Sebagai salah satu syarat kelulusan Program Studi Strata Satu (S1) Informatika di Universitas Amikom Yogyakarta.

1.6 Manfaat Penelitian

Dalam pembuatan penelitian ini, adapun maksud adalah sebagai berikut :

1. Pemrosesan klasifikasi pada deteksi serangan DDoS menjadi lebih efektif dan efisien setelah dilakukan seleksi atribut.
2. Mengetahui atribut atau fitur apa saja yang relevan pada dataset.
3. Mengetahui pengaruh penerapan seleksi atribut pada kinerja algoritma klasifikasi

1.6 Metode Penelitian

1.6.1 Metode Pengumpulan Data

Metode pengumpulan data untuk training dan test dalam penelitian ini didapatkan dari repository milik University of New Brunswick bagian DDoS Evalution Dataset (CIC-DDoS2019). Peneliti mengumpulkan data sebagai referensi yang didapat di berbagai media seperti internet, e-book, atau jurnal online yang berkaitan dengan penelitian.

1.6.2 Tahap-tahap Penelitian

Dalam penelitian ini, penelitian ini menggunakan dataset yang berasal dari repository milik University of New Brunswick bagian DDoS Evalution Dataset (CIC-DDoS2019). Dataset yang diambil yaitu 20% pada dataset jenis serangan *Portmap*. Data tersebut dilakukan preprocessing dengan cara menghapus label yang terdapat jenis baris data dan missing value. Kemudian dataset dilakukan seleksi fitur yang memiliki tingkat relevan yang tinggi satu sama lain, sedangkan fitur yang memiliki relevansi rendah atau tidak ada sama sekali maka tidak dipakai. Data hasil seleksi fitur digunakan untuk menguji algoritma klasifikasi J48, Naive Bayes dan RBF Network. Sehingga peneliti bisa membandingkan tingkat kinerja masing-masing algoritma klasifikasi.

1.6.3 Sistematika Penulisan

Pada bagian ini dituliskan urutan-urutan dan sistematika penulisan yang dilakukan. Berikut ringkasan mengenai isi masing-masing bab:

1. **BAB I PENDAHULUAN**

Pada bab ini dijelaskan mengenai latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitan, dan sistematika penulisan dalam penelitan.

2. **BAB II LANDASAN TEORI**

Pada bab ini akan dijelaskan mengenai landasan teori-teori dan kajian pustaka dari berbagai penelitian yang memiliki keterkaitan dengan penelitian ini. Hal tersebut berguna untuk memperkuat dasar, analisa, penulisan dan alasan dilakukannya penelitian ini. Sumber dari landasan teori ini juga berasal dari buku, jurnal yang secara fisik maupun berasal dari internet.

3. **BAB III METODOLOGI PENELITIAN**

Pada bab ini akan dijelaskan mengenai langkah-langkah penelitian beserta metode yang digunakan.

4. **BAB IV HASIL DAN PEMBAHASAN**

Pada bab ini akan dilakukan perancangan sistem serta pembahasannya. Kemudian di bab ini juga hasil penelitian akan didapatkan dan dilakukan pembahasannya.

5. **BAB V KESIMPULAN**

Pada bab ini berisi kesimpulan dari penelitian ini dan juga saran bagi penelitian mendatang yang berasal dari kekurangan dari penelitian ini.