BAB 1 PENDAHULUAN

1.1 Latar Belakang

Keamanan jaringan merupakan aspek penting dalam bidang teknologi informasi saat ini. Semakin banyak pengguna dan semakin luas jangkauan komunikasi, maka semakin banyak pula peluang serangan. Sebagai gambaran, pada survey yang dilakukan oleh Lab Kaspersky 2017, 33% organisasi mengalami serangan *DDoS* pada tahun 2017, dibandingkan dengan 17% di tahun 2016. Dari organisasi yang terkena serangan *DDoS*, 20% adalah bisnis yang sangat kecil, 33% adalah UKM, dan 41% adalah perusahaan. Serangan atau *Intrusion* dapat diartikan sebagai sebuah tindakan untuk memasuki wilayah atau akses yang tidak sah yang dapat membahayakan perangkat jaringan lainnya. (William Stallings, 2005).

Untuk melakukan pendeteksian atau pencegahan berbagai potensi serangan telah dikembangkan suatu sistem atau metode yang dikenal dengan *Intrusion Detection System (IDS)*. IDS adalah sebuah perangkat lunak atau perangkat keras yang dapat mendeteksi aktivitas yang mencurigakan dalam sebuah sistem atau jaringan (Monika Kusumawati, 2010). IDS memiliki dua metode dalam melakukan pendeteksian yaitu *Rule Based (Signature Based) dan Behavior Based.* Pendeteksian berbasis *Signature Based* dilakukan dengan mencocokkan lalu lintas jaringan dengan sebuah *rules* yang dibuat oleh administrator dan disimpan dalam sebuah *database*. Pendeteksian jenis ini membutuhkan pembaruan terhadap *database rule* pada *IDS* yang bersangkutan. Berbeda halnya dengan *Behavior Based* yang mendeteksi serangan dengan membandingkan pola dari sebuah dataset menggunakan sebuah metode untuk proses klasifikasi.

Secara umum *Behavior based* dalam proses kerjanya melakukan perbandingan pola atau aktivitas yang ada pada sebuah data, kemudian dilakukan klasifikasi dengan sebuah metode dan menghasilkan sebuah model. Dari model yang sudah dibangun tersebut diuji dengan data *testing* menghasilkan sebuah output untuk melihat akurasi apakah sebuah *traffic* yang ada dapat dikategorikan sebagai intrusi atau bukan. Maka dalam hal ini dibutuhkan sebuah metode yang digunakan untuk proses klasifikasi untuk menghasilkan akurasi yang akurat.

Terdapat beberapa penelitian mengenai perbandingan metode klasifikasi, yang pertama penelitian oleh Srinivas Mukkamala dan Andrew H. Sung (2003). Penelitian ini membahas tentang seleksi fitur yang digunakan untuk IDS dengan menggunakan metode ANN dan SVM agar IDS mencapai performa maksimal dengan menggunakan dataset DARPA 1998. Menurutnya bahwa kinerja algoritma SVM lebih baik jika dibandingkan dengan ANN dalam hal solusi yang dicapai untuk kasus pengklasifikasian IDS. Kemudian penelitian kedua yang dilakukan oleh Mrutyunjaya Panda dan Mana R. Patra (2007) yang menerapkan metode *Naive Bayes* pada deteksi intrusi berbasis anomali menggunakan dataset KDDCup'99. Penelitian ini menyatakan bahwa kinerja algoritma *Bayesian* lebih efisien dalam mengklasifikasikan *Network* IDS (NIDS) dibandingkan ANN. Selanjutnya penelitian

yang dilakukan Dwi Widiastuti (2012) yang melakukan perbandingan antara algoritma SVM, *Naive Bayes* dan *Decision Tree* dalam melakukan klasifikasi serangan pada sistem deteksi intrusi dimana data yang digunakan yaitu KDD Cup'99. Penelitian ini menyatakan kinerja algoritma *decision tree* lebih baik dibandingkan dengan algoritma SVM dan NBC. Dari ketiga penelitian tersebut, dataset yang digunakan merupakan dataset lama sehingga dibutuhkan sebuah dataset yang lebih baru untuk deteksi serangan saat ini. Metode klasifikasi yang digunakan juga belum ada yang membandingkan secara spesifik nilai akurasi dari metode *Naive Bayes* dan *Support Vector Machine (SVM)* dengan menggunakan kernel *Linear*, *Polynomial* maupun *Sigmoid*.

Dari penjabaran tersebut, maka dilakukan sebuah penelitian yang berjudul "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode *Naive Bayes* dan *Support Vector Machine (SVM)*". Tujuan dari penelitian ini untuk melihat tingkat akurasi dari metode *Naive Bayes*, SVM *Linear*, SVM *Polynomial*, dan SVM *Sigmoid* dengan menggunakan dataset ISCX 2012. Dataset ini dipilih karena merupakan dataset baru yang dikembangkan dari tahun 2009 sampai 2012 oleh Fakultas Ilmu Komputer, *Universitas New Brunswick*. Dimana pada penelitian sebelumnya dataset yang digunakan yaitu KDD Cup 99 dan DARPA 98 yang merupakan dataset lama sehingga kurang akurat jika dilakukan pengujian deteksi serangan saat ini. Selain itu, Dataset KDD Cup 99 yang merupakan hasil preprocessing data mentah dari DARPA 98 yang memiliki kelemahan distribusi data intrusi dan data normal yang tidak natural yaitu sekitar 80% dari dataset tersebut merupakan data intrusi. Dengan demikian, performansi suatu teknik pendeteksi intrusi dengan distribusi data seperti ini tidak dapat menggambarkan akurasi jumlah *false positive* yang sebenarnya. Terdapat beberapa tahapan dalam penelitian ini, tahap pertama dataset ISCX dilakukan *preprocess*, selanjutnya menghilangkan beberapa fitur untuk proses klasifikasi, setelah selesai data siap dimasukkan dalam *classifier*. Proses pengujian memanfaatkan pengambilan sampel dengan teknik *random sampling* dimana persentase data *training* 60% data *testing* 40% sehingga menghasilkan persentase model akurasi serta *output* berupa *confusion matrix* dan kurva *ROC (Receiver Operating Characteristic).*

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan, maka dapat dirumuskan beberapa permasalahan sebagai berikut:

1. Bagaimana tahapan klasifikasi serangan menggunakan metode *behavior based*?
2. Bagaimana mekanisme pengolahan dataset ISCX menjadi data yang siap dimasukkan *classifier*?
3. Fitur apa saja yang digunakan untuk melakukan klasifikasi?

4. Bagaimana performa dari *confusion matrix* dan *ROC (Receiver Operating Characteristic)*?

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk menganalisis perbandingan akurasi deteksi serangan pada jaringan komputer dengan menggunakan metode *Naive Bayes*, SVM *Linear*, SVM *Polynomial* dan SVM *Sigmoid*.

1.4 Manfaat

Manfaat yang diperoleh penulis adalah penulis mendapat pengetahuan mengenai metode yang terbaik antara SVM dan *Naïve Bayes* serta mengetahui fitur untuk mendapatkan akurasi yang tinggi. Untuk Program Studi Teknik Informatika, Jurusan Teknik Informatika serta Fakultas Ilmu Komputer, penelitian ini dapat menambah referensi penelitian di bidang jaringan khususnya mengenai *Intrusion Detection System*, *Naive Bayes*, SVM *Linear*, SVM *Polynomial*, SVM *Sigmoid* serta konsep deteksi sistem.

1.5 Batasan Masalah

Adapun batasan masalah pada tugas akhir ini, yaitu:

a. Metode yang digunakan untuk melakukan klasifikasi yaitu *Naive Bayes,* SVM *Linear*, SVM *Polynomial*, dan SVM *Sigmoid.*
b. *Library* yang digunakan yaitu *scikit-learn.*
c. Dataset yang digunakan merupakan data dari ISCX 2012 pada tanggal 14 Juni 2012.
d. Penelitian ini hanya melakukan pengujian tidak melakukan deteksi.
e. Hasil keluaran hanya berupa nilai *accuracy, precision, recall, f1 score* dari *confusion matrix* dan kurva *ROC (Receiver Operating Characteristic).*

1.6 Sistematika Pembahasan

Sistematika pembahasan dari penyusunan penelitian yang direncanakan adalah sebagai berikut:

BAB I PENDAHULUAN

Pendahuluan terdiri dari latar belakang, identifikasi dan pembatasan masalah, rumusan masalah, tujuan dan manfaat serta sistematika pembahasan dari penelitian ini.

BAB II LANDASAN KEPUSTAKAAN

Bab landasan kepustakaan menjelaskan tentang teori-teori yang digunakan dalam penelitian, temuan dan bahan penelitian terdahulu yang diperoleh dari beberapa refrensi yang menunjang penelitian dalam penulisan skripsi.

BAB III METODOLOGI

Bab metodologi menjabarkan tahapan-tahapan dalam melakukan penelitian. Tahapan-tahapan seperti studi literatur, pengumpulan data, perancangan lingkungan pengujian, implementasi lingkungan pengujian, serta analisa hasil dibahas secara umum.

BAB IV SIMULASI

Bab simulasi memuat tentang perancangan lingkungan pengujian serta implementasi lingkungan pengujian dalam melakukan proses klasifikasi dataset menggunakan metode *Naive Bayes* dan SVM.

BAB V HASIL DAN PEMBAHASAN

Bab hasil dan pembahasan memuat mengenai hasil dari metode klasifikasi yang dilakukan pada data sekunder dimana data yang didapat pada ISCX *testbed* 14 Juni 2012 yaitu *confusion matrix* dan juga kurva ROC. Serta akan dilakukan analisa untuk mengetahui performa nilai akurasi pada masing-masing metode.

BAB VI PENUTUP

Pada bab penutup memuat kesimpulan dari keseluruhan penelitian. Kesimpulan ini dibuat setelah melalui beberapa proses dalam melakukan klasifikasi data. Saran-saran juga diberikan agar hasil dari penelitian ini dapat diperbaiki dan disempurnakan apabila penelitian ini dikembangkan kemudian.