1. Pendahuluan

1.1 Latar Belakang Masalah

Jaringan komputer sering menjadi target para *hacker* dan *intruder*. Dalam melakukan serangan, para *hacker* atau *intruder* melakukan aktivitas diluar aktivitas normal jaringan. Aktivitas inilah yang disebut sebagai intrusi.

Intrusion detection adalah proses untuk memonitor *event* yang terjadi pada jaringan komputer dan melakukan analisis data tersebut untuk mengetahui adanya intrusi. *Intrusion Detection System* (IDS) memainkan peranan penting dalam menjaga *integrity, confidentiality dan availability* dari sumber daya jaringan komputer [2]. Tujuan utama IDS adalah mengklasifikasikan aktivitas jaringan apakah termasuk aktivitas normal atau intrusi.

Ada 2 pendekatan dalam mengenali intrusi dalam IDS, yaitu *anomaly* dan *misuse detection*. *Anomaly detection* adalah mengidentifikasi perilaku tak lazim yang terjadi (*anomaly*) dalam *host* atau *network*. Detektor berfungsi dengan asumsi bahwa intrusi itu berbeda dengan aktivitas normal. Oleh karena itu dalam *anomaly detection* menggunakan *clustering* (pengelompokan) supaya jika ada aktivitas yang berbeda, segera dicurigai sebagai intrusi. *Anomaly detection* menggunakan pendekatan *unsupervised clustering*, yang mampu mendeteksi intrusi tanpa harus mempelajari data terlebih dahulu . Sedangkan *misuse detection* termasuk *supervised learning*, dimana dibutuhkan data sebelumnya sebagai pembelajar. Bedasarkan data yang sudah dipelajari tersebut akan dibentuk *pattern* (pola) untuk masing-masing kelas. Dengan demikian, metode klasifikasi dapat menangani *misuse detection*.

Yang menjadi fokus dalam penelitian tugas akhir ini adalah *misuse detection* yaitu menggunakan metode klasifikasi. *Misuse detection* adalah menganalisis aktivitas sistem, mencari *event* yang cocok dengan pola perilaku yang dikenali sebagai serangan . Salah satu kelebihan dari *misuse detector* adalah mampu dengan cepat dan handal mendiagnosa penggunaan teknik serangan tertentu. Kekurangan dari *misuse detection* ini adalah banyak *false positive* (terdeteksi ada serangan padahal tidak terjadi serangan). Telah banyak metode yang digunakan untuk *misuse detection* antara lain: ANFIS [21], Hidden Markov Model [19], SVM, Naive Bayes [16], C4.5 [14].

Proses klasifikasi *misuse detection* menggunakan *dataset* KDD Cup 1999. Karakteristik dataset ini adalah *dataset* yang memiliki banyak *features* (42 features) yang terdiri dari *features* kategoris dan numerik. Selain itu *dataset* KDD Cup 1999 tergolong *unbalanced data*, dimana ada kelas yang terdistribusi tidak seimbang diantara kelas yang berbeda [13].

Melihat karakteristik dataset KDD Cup 1999, diperlukan metode yang mampu menangani masalah *unbalanced data* dan menangani banyak *features* yang berbeda-beda jenisnya (kategoris dan numerik).

Random Forest (RF) merupakan salah satu metode yang digunakan untuk klasifikasi dengan membangun banyak pohon klasifikasi. RF dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap *node* (simpul diatasnya) dan diakumulasikan hasil klasifikasi dari

setiap pohon, kemudian dipilih hasil klasifikasi yang paling banyak muncul [22] . Banyaknya pohon yang akan dibentuk sangat berpengaruh terhadap tingkat akurasi hasil klasifikasi. Semakin banyak pohon, semakin akurat hasil klasifikasinya. Selain itu juga RF dapat menangani input variabel yang besar, menyeimbangkan *error* dalam *unbalanced dataset*. Untuk menangani *unbalanced data*, algoritma RF mengalami sedikit modifikasi pada pemilihan data *training*, yaitu dengan menyeimbangkan jumlah *record* pada kelas mayor dan minor. Teknik ini disebut *Balanced Random Forest* (BRF).

Dalam algoritma RF, diperlukan algoritma untuk membangun *tree*. Salah satu algoritma yang dapat digunakan adalah algoritma *Classification and Regression Tree* (CART). CART membagi pohon keputusan dengan teknik *binary tree* dan dapat diterapkan pada variabel numerik dan kategoris sekaligus [2]. Telah dibuktikan pada [3], bahwa CART cocok diterapkan untuk data dengan variabel yang banyak dan kompleks.

Dalam penelitian ini dilakukan penggabungan metode RF dan CART. Dengan kekuatan metode – metode yang digunakan dalam penelitian ini, diharapkan dapat menjawab permasalahan *misuse detection* menggunakan dataset KDD Cup 1999. Permasalahan *unbalanced data* ditangani oleh metode RF dan permasalahan perbedaan jenis *features* ditangani oleh CART.

1.2 Perumusan Masalah

Permasalahan yang menjadi objek dari penelitian tugas akhir ini, terdiri atas :

a. Bagaimana cara menerapkan algoritma Random Forest dan CART untuk membangun model klasifikasi pada *misuse* IDS?
b. Bagaimana performansi (bedasarkan *precison, recall, F Measure*) sistem yang dibangun dengan algoritma RF dan CART dalam mendeteksi intrusi (diklasifikasikan menjadi Normal, Probe, DoS, U2R, R2L)?
c. Berapa banyak jumlah pohon optimum yang dibentuk agar dapat menghasilkan tingkat akurasi yang lebih baik dalam mendeteksi intrusi (diklasifikasikan menjadi normal, probe, DoS, U2R, R2L)?

Hipotesa awal dari penelitian ini adalah gabungan metode RF dan CART dapat menghasilkan akurasi yang lebih baik daripada *single classifier* untuk klasifikasi dalam mendeteksi intrusi.

Batasan masalah untuk penelitian ini adalah:

1. *Preprocessing* (Normalisasi, *remove outlier*, *feature selection*) dilakukan menggunakan tools Weka 3.6.1
2. *Dataset* yang digunakan berasal dari data KDD Cup 1999
3. Sistem yang dibangun tidak *real-time*

1.3 Tujuan

Tujuan dengan dilakukannya penelitian ini adalah:

1. Menganalisis dan mengimplementasikan algoritma RF dan CART untuk membangun model yang dapat digunakan untuk klasifikasi dalam *misuse detection* IDS bedasarkan *dataset* KDD Cup 1999
2. Menganalisis performansi sistem yang dibangun (berdasarkan *precison, recall, F Measure*) serta faktor yang mempengaruhi keakuratan dari sistem yang dibangun
3. Menganalisis pengaruh banyaknya pohon yang dibentuk terhadap tingkat akurasi dalam mendeteksi intrusi (diklasifikasikan menjadi Normal, Probe, DoS, U2R, R2L)

1.4 Metodologi Penyelesaian Masalah

Metodologi yang dilakukan untuk menyelesaikan permasalahan adalah sebagai berikut:

1. Studi literatur

   Melakukan pencarian serta mempelajari informasi dan pembelajaran tentang *misuse* IDS, khususnya mengenai konsep dan cara kerja metode RF dan CART untuk klasifikasi *misuse* IDS. Selain itu juga mempelajari karakteristik dan cara *preprocessing data* yang dapat diterapkan pada *dataset* KDD Cup 1999.
2. Pengumpulan data-data

   Melakukan pencarian data yang digunakan untuk penelitian Tugas Akhir ini. Data yang dicari adalah *dataset* KDD Cup 1999, serta keterangan *features* didalamnya.
3. Analisis modifikasi CART

   Melakukan analisis kemungkinan modifikasi atau pengembangan terhadap metode CART yang akan digabungkan dengan metode RF, bedasarkan kelebihan dan kekurangan dari metode RF dan CART.
4. Analisis dan perancangan aplikasi

   Melakukan analisis dan perancangan aplikasi yang akan dibentuk menggunakan metode RF dan CART sehingga dapat digunakan untuk menghitung akurasi dari klasifikasi *misuse* IDS.
5. Implementasi aplikasi

   Melakukan implementasi aplikasi sesuai dengan hasil analisis dan perancangan metode RF dan CART sehingga dapat digunakan untuk menghitung akurasi dari klasifikasi *misuse* IDS.
6. Pengujian aplikasi

   Melakukan pengujian aplikasi dan menganalisis hasil keluaran aplikasi, sejauh mana keakuratan dari *detection model* yang dibangun dengan menggunakan metode RF dan CART dalam mengklasifikasikan *data testing* KDDCup 1999.
7. Pembuatan laporan Tugas Akhir

   Melakukan penyusunan laporan hasil penelitian yang telah dilakukan serta memberikan kesimpulan dari hasil penelitian tersebut.