1. Pendahuluan

Latar Belakang

Pengguna internet di berbagai negara terus bertumbuh dari tahun ke tahun. Menurut survey Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) pada tahun 2023 sebanyak 78,19 atau sekitar 215 juta penduduk Indonesia telah menjadi pengguna aktif internet. Dengan meningkatnya pengguna internet saat ini, keamanan jaringan menjadi hal yang penting untuk mengamankan setiap data. Karena semakin berkembangnya teknologi dan informasi, maka serangan yang dilakukan oleh pihak yang tidak diinginkan akan semakin meningkat. Sebagai contoh, pada tahun 2020 situs web DPR RI mengalami *down* dan berganti nama, setelah diselidiki ternyata ini akibat dari serangan DoS dari peretas. Untuk meningkatkan keamanan dan pencegahan terjadinya serangan oleh pihak yang dapat merugikan, maka dibutuhkan suatu sistem yang mampu menanganinya. Salah satu sistem yang dapat digunakan untuk mencegah resiko serangan tersebut adalah *Intrusion Detection System* (IDS). Berdasarkan strategi pendeteksiannya, ada dua jenis deteksi intrusi, yaitu *signature-based detection* dan *anomaly-based detection*.

Signature-based IDS (SIDS) adalah suatu pendeteksian dengan merancang terlebih dahulu serangan dan pola serangan dan pola serangan kemudian dikumpulkan menjadi suatu data klasifikasi [13]. Data audit yang dikumpulkan oleh IDS akan dibandingkan dengan data klasifikasi, apabila kedua data tersebut memiliki kesamaan maka akan muncul peringatan bahwa ada serangan, apabila data tidak cocok dengan data klasifikasi serangan, maka itu akan termasuk aktivitas yang aman. SIDS ini mampu mendeteksi serangan dengan baik dan menghasilkan *false positive rate* yang rendah karena dia merancang terlebih dahulu aturan untuk setiap serangan, namun kekurangannya adalah harus merancang aturan dengan manual dan tidak bisa mendeteksi serangan yang belum dikenali atau *zero day attack* [16]. Sedangkan *anomaly-based* IDS (AIDS) adalah suatu pendeteksian dengan merancang lalu lintas normal dari jaringan, kemudian mencari aktivitas anomali yang tidak sesuai dengan model yang dirancang [19]. Pendeteksian ini mengasumsikan bahwa semua aktivitas yang tidak normal akan dikategorikan sebagai serangan. Kelebihan dari metode ini adalah mampu mendeteksi intrusi serangan baru karena menganggap semua aktivitas pada jaringan yang menyimpang akan dikategorikan sebagai serangan.

Banyak penelitian yang telah dilakukan dalam rangka menemukan metode klasifikasi terbaik untuk AIDS, salah satunya adalah penelitian yang dilakukan oleh Farhat et al [8], mereka melakukan uji coba pada dataset NSL-KDD dengan beberapa metode klasifikasi *supervised learning* dan *unsupervised learning*. Pada hasil peneli-tiannya ditunjukan bahwa *Random Forest* merupakan algoritma terbaik dengan *True Positive Rate* (TPR) sebesar98,6% dan *False Positive Rate* (FPR) sebesar 1,5%. Pada dataset yang sama, Nathiya et al [14] melakukan pe- nelitian dengan membandingkan 3 metode klasifikasi yaitu *Support Vector Machine* (SVM), Naive Bayes, dan *Decision Tree* (J48) pada *tools* WEKA. Penelitiannya menunjukan bahwa *Decision Tree* mempunyai performansilebih baik daripada 2 algoritma lainnya dengan 99,3% akurasi, 99% TPR dan 0,5% FPR. Sedangkan pada pene- litian Sivanantham et al [20], mereka membandingkan beberapa algoritma *classifier* dengan menggunakan teknik *Boosting* pada WEKA. Mereka menggunakan *AdaptiveM1 Boosting* atau AdaBoost untuk *Boosting classifier* pada dataset NSL-KDD. Berdasarkan hasil penelitian ini, *AdaBoost* dipadukan dengan *Decision Tree* mendapatkan skor paling tinggi dengan nilai akurasi sebesar 98,45% dan FPR 1,59%, mereka mengungkapkan bahwa dengan teknik *Boosting* akan meningkatkan performansi algoritma.

Berdasarkan penelitian yang dikaji sebelumnya, terdapat perbedaan pendapat dari beberapa penelitian yang menyebutkan hasil algoritma klasifikasi mereka meskipun menggunakan dataset yang sama dan dengan *tools* yang sama. Beberapa faktor yang menyebabkan perbedaan tersebut dari versi WEKA yang digunakan, perbedaan *device* yang digunakan, pembagian data latih dan data uji serta metode klasifikasi yang diuji. Maka perlu dilakukan penelitian untuk membuktikan metode klasifikasi manakah yang terbaik untuk membangun AIDS menggunakan WEKA.

Topik dan Batasannya

Berdasarkan latar belakang diatas, tugas akhir ini memiliki rumusan masalah menguji metode klasifikasi manakah yang memiliki performansi yang paling baik dari 3 metode klasifikasi yang telah diuraikan pada penelitian-penelitian sebelumnya. Selain itu, rumusan masalah lain yang dapat diangkat adalah membuktikan bahwa apakah dengan teknik *Boosting* mampu meningkatkan performansi suatu algoritma *classifier*. Tugas akhir ini memiliki beberapa batasan masalah sebagai berikut:

1. Pengujian hanya dilakukan pada dataset NSL-KDD dan UNSW-NB15.
2. Pengujian akan dilakukan pada *tools* WEKA.
3. Metrik uji yang digunakan yaitu *accuracy*, *precission*, *recall*, *f1-score* dan waktu uji.
4. Model algoritma yang digunakan yaitu *Random Forest*, J48 dan J48 menggunakan AdaBoost.

Tujuan

Tujuan dari tugas akhir ini adalah untuk melakukan pengujian terhadap 3 model Algoritma yaitu *Random Forest*, J48 dan J48 menggunakan *AdaBoost* untuk mengetahui model algoritma mana yang mempunyai performansi terbaik untuk AIDS. Penelitian ini juga menguji apakah performansi dari algoritma J48 akan meningkat apabila dikombinasikan dengan *AdaBoost*. Selain itu, penelitian ini menguji apakah AIDS mampu menangani *zero-day attack*.

Organisasi Tulisan

Pada tugas akhir ini terdapat 5 bab. Bab pertama membahas latar belakang, topik dan batasan penelitiandan tujuan penelitian. Bab kedua berisi tentang studi terkait referensi dalam pengerjaan tugas akhir. Bab ketiga menjelaskan sistem yang akan dibangun. Bab keempat membahas tentang evaluasi dan analisis terhadap penelitian yang dilakukan. Bab kelima membahas kesimpulan dan saran dari penelitian.