BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam Perkembangan teknologi,internet memiliki peran yang sangat besar untuk mendukung kinerja dan aktivitas dalam kehidupan sehari-hari,sehingga tidak dapat di pungkiri pengunaan intenet semakin meningkat,menurut data hasil Survei APJII(Asosiasi penyelenggara jasa internet Indonesia) dari jumlah 210 juta jiwa penguna internet di tahun 2022, mengalami peningkatan sebesar 1,17% di tahun 2023 dengan total jumlah 215 juta jiwa dari total populasi 275 juta jiwa penguna internet di Indonesia(APJII 2023), Dengan semakin banyaknya penguna internet tidak menutup kemungkinan rentan akan terkena sasaran serangan kejahatan dalam dunia Cyber,pada umumnya focus utama dari serangan jaringan adalah untuk melakukan Tindakan ancaman terhadap bisnis komersial dan kehidupan sehari-hari,kerugian yang di dapat bisa berupa pencurian data, rusaknya system,Sniffing,Virus,dan sebagainya.Maka dari itu salah Satu Solusi untuk Mencegah terjadinya serangan jaringan dapat melakukan Tindakan berupa IDS (Intrusion Detection System).

IDS (Intrusion Detection System) Merupakan sistem yang memonitor lalu lintas jaringan untuk aktivitas yang meragukan dan memberikan peringatan ketika aktivitas tersebut terungkap(Rama Devi and Abualkibash 2019). Jadi tujuan dari IDS untuk mendeteksi penyalahgunaan dalam jaringan yang tidak sah.oleh karena itu,system IDS menjadi alat penting dalam jaringan computer agar dapat menyediakan lingkungan jaringan yang lebih aman.menurut penelitian terdahulu penelitian dapat dilakukan dengan teknik machine learning.

Machine learning dalam IDS(Intrusion detection system) yaitu melakukan pendeteksian apakah ada penyusupan atau tidak dalam aktifitas jaringan tersebut yang Dimana aktifitas jaringan tersebut memiliki jumlah data yang berjumlah ribuan hingga jutaan data,kemudian data tersebut diolah menggunakan beberapa algoritma machine learning seperti Random Forest, Decision Tree, Naive Baiyes, gradient boost,K-Nearest

Neighbor(KNN) Dan lain-lain, yang Dimana nantinya akan melalui tahapan - tahapan seperti feature selection,transform data, cleansing data, dan lain-lain. Sehingga mendaptakan hasil akurasi dan waktu pengujian dan test data dengan baik.

Pada peneleitian ini untuk mendeteksi aktifitas jaringan, model yang digunakan yaitu superviced learning dengan algotitma *Random forest, Decision Tree, dan XG-Boost,* ke-3 model tersebut tergolong sudah sering digunakan dalam penelitian-penelitian sebelumnya, dikarenakan model tersebut mampu bekerja dalam jumlah record data yang banyak dengan efektif. *Random Forest* memiliki keunggulan yang dimana dapat bekerja dengan data yang besar, fleksibel untuk digunakan serta tidak sensitive terhadap outlier. *XG-Boost* memiliki keunggulan diantaranya mampu bekerja dalam jumlah data yang banyak, memiliki kemampuan pemrosesan paralel,fleksibel dalam penggunaan,serta memiliki kinerja yang tinggi sehingga dapat memprediksi secara objektif dan akurat (Erkamim et al. 2023). Sedangkan model *Decision Tree* memiliki keunggulan yang dimana dapat di interpretasikan oleh manusia dengan mudah (Devia 2023).

Penelitian tentang Intrusion Detection System sudah Banyak dilakukan sebelumnya dengan berbagai macam Algoritma Machine learning maupun deep learning,penelitian yang dilakukan oleh (Tripathy and Behera 2023),peneliitian ini membahas evaluasi kinerja Algoritma machine learning dalam Intrusion Detection system yang Dimana penelitian ini menggunakan Dataset KDD CUP 99,penelitian Ini melakukan evaluasi dengan 15 Algoritma Machine Learning yang Dimana hasil penelitian ini memiliki akurasi tertinggi pada algortima SVM(Support Vector Machine) dengan nilai akurasi 98,08%.Sedangkan untuk model Random Forest mendapatkan akurasi 97,14%, Decision Tree 97,13%, dan untuk model XG-Boost 94,64%. Tidak terlepas dari penelitian tersebut, dirasa performa dari ke-3 model algoritma yang ingin di gunakan oleh peneliti masih bisa ditingkatkan, dikarenakan memiliki banyak keunggulan, sehingga menjadi alasan peniliti mengapa model *Random Forest, Decision Tree, dan XG-Boost* di gunakan

Namun, ada beberapa masalah tertentu, terutama dengan transparansi sistem, di bidang pendeteksian jaringan. Pakar keamanan siber sekarang biasanya mengambil keputusan berdasarkan rekomendasi IDS akibatnya model prediksi harus dimengerti. Kemudian, kompleksitas model menjadi perhatian utama bagi orang-orang saat mereka terlibat, karena model-model ini tidak dapat memberikan informasi apa pun tentang hasil model mereka. Sebagai hasilnya, sangat penting untuk memberikan beberapa informasi terkait dengan prediksi IDS, serta memberikan informasi kepada personel keamanan tentang intrusi yang telah terdeteksi.

Untuk mengatasi masalah ini dan memberikan penjelasan yang lebih baik untuk IDS, sebuah kerangka kerja berdasarkan Shapley Additive Explanations (SHAP) diciptakan. SHAP memiliki landasan teori yang kuat dan dapat diterapkan pada model apapun.yang dimana tujuan dari Shap Values adalah untuk memberikan penjelasan meganai kontribusi fitur terhadap prediksi model.

Dalam penelitian ini akan menggunakan dataset Kdd cup-99,dataset ini berisikan data aktifitas jaringan dalam lingkungan jaringan militer,.dalam penelitian ini akan menggunakan dataset Kdd_cup_10_percent.gz yang Dimana dataset ini di peroleh dari UCI Machine learning.

Berdasrkan penjabaran yang telah dipaparkan sebelumnya maka akan dilakukan Analisis perbandingan dan interpretasi Shap Values dari setiap model algoritma machine learning (Random Forest,Decision Tree, dan XG-boost) untuk menjelaskan performa dalam setiap model algoritma tersebut dan mengetahui kontribusi dari setiap fitur terhadap model algoritma tersbut dalam deteksi intruksi jarigan menggunakan dataset kdd cup 99.

1.2. Rumusan Masalah

Berdasarkan dari latar belakang di atas, dapat dirumuskan permasalahan yang ada dalam penelitian ini yaitu, dapat meningkatkan performa model Random Forest, Decision Tree, dan XG-Boost serta menjelaskan kontribusi dari setiap fitur untuk mengetahui fitur apa saja yang berpengaruh terhadap model algoritma tersebut, ataupun aktifitas jaringan apa saja yang terindikasi serangan dalam dataset aktifitas jaringan *KDD cup-99 10 %* .

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini menggunakan dataset kdd cup 10 persen yang Dimana dataset ini memiliki jumlah data 494.021 dengan jumlah fitur 41
2. Algoritma yang digunakan Random Forest,Decision Tree, dan XG-Boost
3. Interpretasi menggunakan shap values yang di mana memiliki daya komputasi yang besar dikarenakan jumlah data yang juga cukup besar dan untuk melakukan komputasi ini menggunakan Free google colaboratory sehingga untuk visualisasi shap values terbatas.
4. Pada penelitian ini berfokus untuk mengevaluasi kinerja model algortima machine learning dan fitur apa saja yang mempengaruhi model algoritma tersebut.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk memberikan informasi hasil dari evaluasi kinerja machine learning terhadap model Random Forest,Decision Tree,danXg Boost,serta interpretasti dari setiap model machine learning tersebut menggunakan shap value untuk memahami fitur mana yang paling berpengaruh terhadap prediksi dari setiap model algoritma machine learning tersebut. .

1.5. Manfaat Penelitian

Manfaat dari penelitian ini yaitu :

1. Menenerapkan beberapa algoritma machine learning (Random Forest,Decision Tree,Xg-boost) untuk mengklasifikasi normal dan serangan
2. Memberikan informasi penjelasan lebih dalam dengan mengevaluasi kinerja dan akurasi model serta mengeksplorasi fitur mana yang paling berpengaruh terhadap prediksi model dan bagaimana fitur tersebut berinteraksi satu sama lain

1.6. Sistematika Penulisan

Berikut sistematika penulisan dirangkum menjadi lima bab sebagai berikut.

BAB I PENDAHULUAN

Pada bab ini menjelaskan mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI

Pada bab ini menjelaskan mengenai tinjauan pustaka dan landasan teori yang dijadikan sebagai acuan oleh penulis dalam melakukan penelitian yang berkaitan dengan topik penelitian ini.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan tahapan dalam tugas akhir yang akan dilakukan dan kebutuhan apa saja yang di perlukan dalam penelitian ini.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai pengolahan dataset hingga hasil yang telah diperoleh dari penelitian ini.

BAB V KESIMPULAN

Bab ini merupakan penutup yang berisi kesimpulan berdasarkan hasil dari penelitian yang telah dilakukan.