1. Pendahuluan

Latar Belakang

Dengan perkembangan Internet dan *Artificial Intelligence* (AI), orang-orang menikmati kenyamanan luar biasa dari Internet yang terus-menerus terancam oleh serangan siber pada saat yang bersamaan[1]. Dengan memahami kelemahan jaringan sangat penting untuk mendeteksi dan memitigasi serangan siber secara efektif. Di antara berbagai jenis ancaman siber, *Advanced Persistent Threat* (APT) menimbulkan tantangan yang sangat signifikan[2].

Serangan APT adalah serangan terkonsentrasi dan terarah yang dirancang khusus untuk setiap target tertentu dan korban tertentu untuk mencari informasi berharga dan menyebarkannya ke luar[3]. Metode serangan ini sering digunakan untuk menyerang hal penting seperti lembaga pemerintah, organisasi, dan perusahaan. Karakteristik, proses, dan siklus hidup serangan APT sering kali mencakup memata-matai, menyerang, meningkatkan hak istimewa, mencuri informasi, dan membersihkan jejak[4]. Serangan APT berbeda dari serangan siber lainnya dalam hal tingkat kesulitan, tujuan, metode dan sumber daya yang digunakan. Serangan ini memerlukan pendekatan yang lebih aktif dan berkelanjutan untuk mendeteksi dan menangani ancaman tersebut. Urgensi dari pendeteksian serangan APT adalah mendeteksi seawal mungkin sebelum serangan lebih lanjut masuk kedalam jaringan.

Dalam sebuah insiden pada tahun 2021, infeksi *spyware* yang mencurigakan diidentifikasi dalam lalu lintas data di mana *spyware* tersebut menunggu selama bertahun-tahun untuk mencuri data dari jaringan organisasi target. Kemudian *McAfee* memeriksa masalah ini selama 2 bulan dan menemukan dua serangan APT yang kemungkinan besar berasal dari Tiongkok. Kasus serangan APT lainnya, yaitu *Deep Panda*, menargetkan tenaga kerja Badan Intelijen Amerika Serikat untuk mengumpulkan informasi sensitif pada tahun 2015, di mana ditemukan bahwa peretas Tiongkok berada di balik serangan tersebut. Para penyerang mengeksploitasi kode untuk malware *Deep Panda* untuk mengkompromikan informasi pribadi sekitar empat juta personel Amerika Serikat[5].

Beberapa cara untuk mendeteksi APT telah ada seperti pada penelitian Do Xuan dan kawan-kawan[6] [7]yang menggunakan *machine learning* pada penelitian ini didapatkan hasil akurasi 97.56 % dan 99,98% serta pada penelitian Dao dan kelompoknya[3] metode yang digunakan yaitu *deep learning* pada penelitian ini didapatkan hasil akurasi yaitu 98,72% . Terdapat juga penelitian Saini dan kelompoknya[5] yang menggunakan *hybrid ensemble machine learning* untuk mendeteksi serangan APT dan mendapatkan hasil akurasi 98,92%. Dalam beberapa penelitian diatas ada beberapa kesamaan yaitu tidak terdapatnya *feature selection* pada model deteksi APT yang menjadi saran solusi pada penelitian ini.

Dalam menggunakan *machine learning* terdapat *feature selection* yang dapat digunakan antara lainnya yaitu, *Univariate Selection, Recursive Feature Elimination, Principal Component Analysis, dan Feature Importance*. *Feature selection* digunakan untuk mengoptimasi dari efisiensi pada tahap training dan testing dan mengurangi nilai *false positive*[8].

Penelitian ini bertujuan untuk mendeteksi serangan APT pada lalu lintas jaringan menggunakan metode *machine learning* yaitu *random forest* algoritma dan dengan *feature selection recursive feature elimination*. Pada studi terkait terdapat hubungan yang signifikan antara algoritma yang digunakan dengan *feature selection* yang digunakan. Dengan adanya *feature selection* efektivas dari suatu algoritma dapat meningkat. Penelitian ini juga memiliki tujuan untuk mendapatkan nilai akurasi dari metode yang digunakan. Dengan demikian, penelitian ini dapat memberikan kontribusi dalam mendeteksi serangan APT pada lalu lintas jaringan yang terus berkembang dimasa yang akan datang.

Rumusan Masalah

1. Bagaimana membuat model serangan *Advanced Persistent Threat* dan model deteksi terhadap serangan *Advanced Persistent Threat?*
2. Bagaimana mengevaluasi model deteksi serangan APT menggunakan metode *machine learning* yaitu algoritma *random forest* dan dengan *feature selection recursive feature elimination?*

Topik dan Batasannya

Topik pada penelitian ini adalah bagaimana performansi dari algoritma *random forest* dengan *feature selection recursive feature elimination* dalam mendeteksi sebuah serangan *Advanced Persistent Threat*. Batasan masalah pada penelitian ini adalah dataset lalu lintas jaringan yang memiliki serangan *Advanced Persistent Threat*. Penelitian ini terbatas menggunakan dataset lalu lintas jaringan yang tersedia memiliki serangan *Advanced Persistent Threat* yang tidak dideteksi secara real-time.

Tujuan

Tujuan dari penelitian ini adalah untuk membuat model serangan *Advanced Persistent Threat* dan model deteksi terhadap serangan *Advanced Persistent Threat*, serta menganalisis akurasi deteksi serangan *Advanced Persistent Threat* menggunakan metode *machine learning* yaitu algoritma *random forest* dan dengan *feature selection recursive feature elimination.*

Hipotesa

Dengan menggunakan metode *random forest* dengan *feature selection recursive feature elimination* dapat menghasilkan nilai akurasi yang lebih baik terhadap deteksi serangan APT.

Organisasi Tulisan

Setelah bagian pendahuluan, penelitian ini akan dilanjutkan ke bagian kedua yaitu studi terkait. Pada bagian studi terkait, akan dibahas tinjauan literatur yang mendukung penelitian ini. Selanjutnya, pada bagian ketiga akan dijelaskan bagaimana sistem yang dibangun, mulai dari tahapan hingga model deteksi yang digunakan. Pada bagian keempat yaitu evaluasi, akan dibahas hasil dari penelitian ini. Terakhir, pada bagian kelima akan dipaparkan kesimpulan dari penelitian yang telah dilakukan.