BAB 1
PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang berkembang pesat seiring dengan kemajuan teknologi informasi, keamanan jaringan menjadi faktor utama dalam menjaga kestabilan dan perlindungan sistem dari berbagai ancaman siber yang semakin canggih dan sulit terdeteksi (Bororing, 2024). Peningkatan lalu lintas data yang signifikan serta semakin kompleksnya infrastruktur jaringan modern menjadikan deteksi serangan sebagai tantangan yang semakin besar, terutama karena metode tradisional sering kali tidak lagi efektif dalam mengenali pola serangan yang terus berkembang (Prabuningrat et al., 2024).

Sebagai solusi utama dalam mengidentifikasi serta mencegah serangan siber, *Intrusion Detection System* (IDS) yang memiliki peran penting dalam mendeteksi aktivitas mencurigakan yang mengindikasikan adanya serangan, IDS memungkinkan respons yang cepat dan efektif untuk mengurangi potensi kerusakan. Namun, efektivitas IDS masih menghadapi tantangan dalam mendeteksi aktivitas mencurigakan di dalam jaringan serta membedakan secara akurat antara lalu lintas jaringan yang normal dan yang berbahaya. Dimana aktivitas normal terkadang terdeteksi sebagai ancaman. Oleh karena itu, diperlukan metode deteksi yang kuat ((Fadhlurrohman et al., 2021).

Salah satu teknik dalam identifikasi jaringan yang normal dan berbahaya menggunakan teknik *clustering* dalam data mining beberapa algoritma populer digunakan dalam proses ini, di antaranya adalah algoritma K-Means dan DBSCAN.

K-Means adalah algoritma *clustering* yang bekerja dengan membagi data ke dalam sejumlah kelompok (*cluster*) berdasarkan jarak terdekat terhadap pusat *cluster* (*centroid*) yang telah ditentukan. Sedangkan DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) merupakan algoritma *clustering* berbasis kepadatan yang mampu mengelompokkan data dengan bentuk *cluster* yang

tidak beraturan serta dapat mendeteksi data yang dianggap sebagai *noise* atau anomali.

Meskipun algoritma ini populer Algoritma K-Means memiliki beberapa keterbatasan dalam menangani data yang memiliki distribusi tidak merata dan kesulitan dalam mendeteksi outlier, tetapi dikenal efisien dalam mengelompokkan data berdasarkan kedekatan antar titik. Sementara itu, terdapat algoritma lain yaitu algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) memiliki keunggulan dalam mengidentifikasi pola klaster yang lebih kompleks dan secara efektif dapat mengenali data yang menyimpang dari kepadatan data utama. Namun, algoritma DBSCAN memiliki kelemahan yang sensitif terhadap parameter eps dan minPts, yang berpengaruh dalam perhitungan kepadatan menetapkan titik yang dapat dijangkau dari yang lain jika terletak dalam jarak tertentu disebut density. Serta kesulitan menangani data dengan kepadatan yang bervariasi (Cinderatama et al., 2022).

Dalam proses pemisahan lalu lintas jaringan normal dan berbahaya, algoritma K-Means dan DBSCAN sering digunakan dalam teknik *clustering*. K-Means bekerja dengan membagi data ke dalam sejumlah *cluster* berdasarkan jarak terdekat ke pusat *cluster* (*centroid*) dan secara bertahap memperbarui posisi pusat tersebut hingga stabil. Algoritma ini efektif untuk data dengan pola terstruktur dan jumlah *cluster* yang telah ditentukan. Sementara itu, DBSCAN mengelompokkan data berdasarkan kepadatan tanpa harus menentukan jumlah *cluster* di awal, serta mampu mendeteksi anomali sebagai *noise*. Dengan karakteristik tersebut, K-Means cocok untuk data yang seragam, sedangkan DBSCAN lebih andal untuk data jaringan yang tidak beraturan dan mengandung lalu lintas berbahaya.

Kombinasi K-Means dan DBSCAN dapat menjadi pendekatan yang baru dalam menganalisis data. Pendekatan ini sudah dilakukan pada penelitian sebelumnya dengan menggunakan data log server.

Pada penelitian ini kombinasi K-Means dan DBSCAN dalam mendeteksi anomali jaringan. K-Means digunakan untuk mengelompokkan data ke dalam *cluster* awal, yang kemudian dianalisis menggunakan DBSCAN untuk mendeteksi outlier atau anomali setiap *cluster*. Pendekatan ini mengurangi kelemahan masing-

masing algoritma dan meningkatkan akurasi dalam mendeteksi anomali jaringan (Irawan, 2024).

Dalam penelitian ini, digunakan Dataset UNSW-NB15 merupakan salah satu dataset representatif dalam pengembangan sistem deteksi intrusi modern karena mencakup beragam jenis serangan lalu lintas jaringan yang lebih realistis, menjadikannya unggul dibandingkan dataset lain seperti NSL-KDD yang cenderung sudah usang dan CICIDS2017 yang memiliki kompleksitas tinggi namun akurasi rendah dalam beberapa pengujian. UNSW-NB15 menawarkan keseimbangan antara fitur numerik dan kategorik serta memiliki distribusi kelas yang lebih, sehingga sangat mendukung penerapan algoritma data mining untuk deteksi anomali secara efektif .

Dataset UNSW-NB15 kumpulan data dalam deteksi serangan dan anomali. Dataset ini terdapat berbagai jenis serangan modern seperti DoS, Exploits, Fuzzers, dan Backdoors, serta lalu lintas normal. UNSW-NB15 memiliki fitur yang mencakup informasi dasar, statistik lalu lintas, serta karakteristik koneksi (Julian, 2024).

Penelitian ini bertujuan untuk kombinasi kinerja K-Means dan DBSCAN dalam mendeteksi anomali jaringan pada IDS yang memiliki kekurangan dan keunggulan . Hasil dari penelitian ini diharapkan dapat memberikan wawasan lebih mendalam mengenai efektivitas kedua metode tersebut dalam meningkatkan deteksi anomali, sehingga dapat berkontribusi terhadap pengembangan sistem keamanan jaringan dalam menghadapi ancaman siber yang terus berkembang.

1.2 Perumusan Masalah

Rumusan masalah dalam pada penelitian ini algoritma K-Means dalam mengelompokkan data tidak memperhitungkan adanya outlier, yang bisa membuat kita perlu mencari metode pengelompokan lain. Adanya false positive, outlier, dan variasi dalam data dapat mempengaruhi efisiensi sistem secara keseluruhan. Penting untuk mencari algoritma pengelompokan lain yang lebih tahan terhadap masalah ini agar hasil pengelompokan menjadi lebih baik.

1.3 Pertanyaan Penelitian

Berikut beberapa pertanyaan yang menjadi dasar dalam melakukan penelitian:

1. Bagaimana penerapan kombinasi k-means dan dbscan dalam mendeteksi ancaman pada *Intrusion Detection System* (IDS)?
2. Bagaimana efektivitas algoritma K-Means dan DBSCAN (*Density-Based Spatial Clustering of Applications with Noise)*dalam menangani false positive ?

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini untuk mengkombinasikan algoritma K-Means dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) dalam mendeteksi ancaman siber pada *Intrusion Detection System* (IDS). Penelitian ini diharapkan dapat memberikan hasil yang efektif dalam kinerja IDS dan keamanan jaringan secara keseluruhan.

1.5 Manfaat Hasil Penelitian

Dengan adanya penelitian ini, diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan hasil kombinasi efektivitas *Intrusion Detection System* (IDS) pada K-Means dan DBSCAN.
2. Dengan adanya kombinasi algoritma ini dapat memberikan kontribusi wawasan algoritma yang lebih efisien.