BAB I
PENDAHULUAN

1.1 Latar Belakang

Keamanan siber menjadi salah satu isu kritis di era digital saat ini, mengingat semakin meningkatnya ketergantungan pada infrastruktur jaringan dalam berbagai aspek kehidupan, mulai dari bisnis, pemerintahan, hingga kebutuhan sehari-hari. Berdasarkan laporan Cybersecurity Ventures, kerugian global akibat serangan siber diperkirakan akan mencapai 10,5 triliun USD pada tahun 2025, meningkat tajam dari 3 triliun USD pada tahun 2015 [1]. Selanjutnya Qrator Labs melaporkan bahwa pada kuartal kedua tahun 2023, serangan DDoS dengan menggunakan metode UDP Flood meningkat dengan persentase 59,31% dari seluruh serangan DDoS yang tercatat [2]. Situasi serupa juga terjadi di Indonesia. Menurut laporan dari SOCRadar, Indonesia mengalami lebih dari 43.000 serangan DDoS sepanjang tahun 2024, dengan metode UDP Flood menjadi salah satu vektor serangan utama. Serangan ini bahkan mencapai puncak intensitas sebesar 693 Gbps, menunjukkan besarnya potensi kerusakan yang dapat ditimbulkan [3].

Di tengah tantangan tersebut, sistem deteksi intrusi berbasis anomali telah menjadi salah satu solusi utama dalam mendeteksi aktivitas mencurigakan pada lalu lintas jaringan. Pemilihan algoritma klasifikasi merupakan komponen penting dalam membangun sistem deteksi serangan yang efektif. Algoritma *machine learning*, seperti K-Nearest Neighbors (KNN), banyak digunakan karena dikenal sebagai algoritma yang sederhana namun efektif, dengan kemampuan adaptasi terhadap berbagai jenis data dan bekerja dengan baik dalam skenario klasifikasi berbasis similaritas [4]. Algoritma K-Nearest Neighbors (KNN) juga telah terbukti efektif dalam tugas deteksi anomali karena kesederhanaannya dan kemampuannya untuk mengklasifikasikan titik data berdasarkan kedekatannya dengan data yang diberi label. Hal ini sangat efektif dalam deteksi anomali jaringan, di mana pola lalu lintas berbahaya sering kali berbeda secara signifikan dari pola lalu lintas normal [5]. Tetapi, performa klasifikasi algoritma KNN dapat terpengaruh oleh pemilihan nilai k

yang tidak optimal. Pemilihan nilai k yang tepat sangat penting dalam algoritma KNN karena nilai k yang berbeda dapat menghasilkan kinerja klasifikasi yang berbeda [6]. Penggunaan nilai k yang tetap dan tunggal dalam tahap pencarian, serta penerapan aturan pemungutan suara mayoritas yang sederhana dalam tahap keputusan, dapat menyebabkan hasil klasifikasi yang kurang akurat dalam beberapa skenario [6].

KNN juga memiliki keterbatasan dalam menangani dataset berskala besar karena perhitungannya harus menghitung kemiripan antara data uji dan seluruh data pelatihan, sehingga meningkatkan waktu komputasi secara signifikan. Selain itu, pada dataset yang tidak seimbang, KNN cenderung mengalami kesulitan dalam mendeteksi kelas minoritas, karena tetangga terdekat dari suatu sampel lebih sering berasal dari kelas yang lebih besar, sehingga sulit untuk mengenali kelas dengan jumlah sampel lebih sedikit [7].

Sebagai pengembangan dari algoritma KNN tradisional, KNN++ merupakan metode KNN yang telah dimodifikasi dengan pendekatan *distance-aware algorithm* guna meningkatkan efisiensi dan akurasi klasifikasi. Algoritma ini mengoptimalkan struktur data dengan menyusun atribut berdasarkan relevansinya terhadap perhitungan jarak, sehingga memungkinkan sistem untuk menentukan tetangga terdekat dengan lebih efisien. Dengan demikian, algoritma KNN++ mampu meningkatkan kinerja deteksi anomali dalam lalu lintas jaringan, terutama dalam sistem deteksi intrusi berbasis anomali [8].

Metode lain yang merupakan pengembangan dari algoritma KNN tradisional adalah CD-KNN, CD-KNN merupakan varian KNN yang dimodifikasi dengan pendekatan dinamis dalam penentuan nilai K untuk setiap instance uji. Metode ini mengelompokkan data latih ke dalam klaster berdasarkan label kelas, lalu menghitung nilai K secara adaptif berdasarkan distribusi kemiripan antar instance dalam tiap klaster. Dengan perhitungan K yang menyesuaikan kepadatan lokal dari tiap klaster, CD-KNN mampu mengurangi risiko bias klasifikasi akibat pemilihan nilai K statis yang tidak optimal. Oleh karena itu, CD-KNN menunjukkan performa yang unggul dibanding KNN konvensional pada sebagian besar dataset uji [9].

Performa algoritma pembelajaran mesin sangat dipengaruhi oleh kualitas data. Untuk meningkatkan akurasi dalam deteksi serangan DDoS, diperlukan seleksi fitur

seperti ANOVA dan penyeimbangan data menggunakan SMOTE (Synthetic Minority Oversampling Technique). Seleksi fitur membantu mengurangi dimensi data dan meningkatkan efisiensi serta generalisasi model [10]. Masalah lain seperti ketidakseimbangan data dapat menyebabkan model bias terhadap kelas mayoritas, menurunkan akurasi deteksi serangan [11]. Untuk mengatasi tantangan ini, SMOTE diimplementasikan dengan menghasilkan sampel sintetis pada kelas minoritas, sehingga distribusi data lebih seimbang dan model lebih sensitif terhadap serangan siber [12].

Penelitian sebelumnya menunjukkan peningkatan efisiensi deteksi intrusi berbasis K-Nearest Neighbor (KNN). Lakshminarayana dan Basarkod (2023) [13] mengusulkan modifikasi KNN berbasis Distance Aware untuk meningkatkan akurasi dan kecepatan klasifikasi pada sistem IDS menggunakan dataset NSL-KDD. Hasilnya menunjukkan kenaikan F1-score sebesar 5,33% dan 3,15% pada dua skenario pengujian. Meski efektif, metode ini masih menghadapi tantangan dalam pemilihan fitur dan ketimpangan data, serta memerlukan validasi lebih lanjut untuk memastikan generalisasi pada dataset lain.

Studi oleh Robindro et al. (2022) mengusulkan Cluster-based Dynamic KNN (CD-KNN) untuk mengatasi kelemahan nilai K yang statis dalam algoritma KNN. CD-KNN menentukan nilai K secara adaptif berdasarkan kepadatan lokal dalam klaster data latih. Metode ini diuji pada sembilan dataset UCI dan menunjukkan peningkatan akurasi hingga 5,1% pada beberapa kasus, terutama pada dataset Parkinson Disease, Sonar, dan Labor. Namun, tantangan masih ada pada dataset berdimensi tinggi dengan data terbatas, serta perlunya eksplorasi lebih lanjut terhadap pengaruh parameter α untuk meningkatkan kestabilan dan generalisasi.

Penelitian ini memberikan pembaruan signifikan dengan cara mengimplentasikan dan membandingkan secara langsung algoritma KNN++ dan CD-KNN menggunakan metode seleksi fitur berbasis ANOVA serta teknik *balancing data* melalui *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan dataset. Validasi menyeluruh dilakukan dengan evaluasi metrik seperti akurasi, presisi, recall, F1-score dan ROC-AUC, sehingga memberikan kontribusi dalam meningkatkan efisiensi dan keandalan algoritma dalam mendeteksi serangan siber

berbasis anomali. Pendekatan ini diharapkan mampu mengatasi keterbatasan yang ada dan memberikan wawasan baru terkait algoritma optimal untuk klasifikasi data siber.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah disajikan, dapat diidentifikasi rumusan masalah yang akan menjadi fokus pembahasan sebagai berikut:

1. Bagaimana cara algoritma *K-Nearest Neighbors*++ dan CD-KNN dapat mengklasifikasikan serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan?
2. Bagaimana hasil kinerja algoritma *K-Nearest Neighbors*++ dan CD-KNN dengan menggunakan seleksi fitur Anova dan tanpa menggunakan seleksi fitur dalam mengklasifikasikan serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan?
3. Bagaimana hasil kinerja algoritma K-Nearest Neighbors++ dan CD-KNN dengan menggunakan seleksi fitur Anova dan tanpa menggunakan seleksi fitur pada data yang dilakukan oversampling SMOTE dalam mengklasifikasikan serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan?

1.3 Tujuan Penelitian

Merujuk dari rumusan masalah diatas, adapun tujuan penelitian yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui cara algoritma *K-Nearest Neighbors*++ dan CD-KNN dalam mengklasifikasikan serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan.
2. Untuk mengetahui hasil klasifikasi serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan menggunakan algoritma *K-Nearest Neighbors*++ dan CD-KNN dengan menggunakan seleksi fitur anova.
3. Untuk mengetahui hasil klasifikasi serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan menggunakan algoritma *K-Nearest*

Neighbors++ dan CD-KNN dengan menggunakan seleksi fitur anova pada data yang telah oversampling SMOTE.

1.4 Manfaat Penelitian

Berdasarkan hasil dari penelitian ini, diharapkan memberikan beberapa manfaat sebagai berikut:

1. Menambah wawasan terkait algoritma *K-Nearest Neighbors*++ dan CD-KNN dalam klasifikasi serangan DDoS, khususnya UDP Flood untuk mengatahui bagaimana algoritma tersebut dapat mengklasifikasikan serangan DDoS UDP Flood berbasis anomali lalu lintas jaringan.
2. Menyumbangkan kontribusi pada pengembangan klasifikasi serangan DDoS UDP *Flood* berbasis anomali lalu lintas jaringan menggunakan algoritma K-Nearest Neighbors++ dan CD-KNN.
3. Hasil penelitian dapat diterapkan dalam proses peningkatan atau pengembangan sistem deteksi intrusi (IDS) dan sistem pencegahan intrusi (IPS).

1.5 Batasan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan pada bagian sebelumnya, terdapat beberapa batasan masalah yang digunakan untuk membatasi cakupan pada penelitian "Klasifikasi Serangan DDoS UDP Flood Menggunakan KNN++ Dengan ANOVA Feature Selection" diantaranya:

1. Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data yang diambil secara tidak langsung yang diambil dari https://www.unb.ca/cic/datasets/ddos-2019.html.
2. Data yang digunakan adalah "DrDos_UDP.csv" yang berasal dari CICDDoS 2019.
3. Pembagian kelas pada dataset berjumlah sebanyak 2 kelas yaitu "DrDos_UDP" dan "*Benign*".
4. Penelitian ini dilakukan dengan menggunakan bahasa python 3 dengan kode editor Google Collaboratory.

5. Keluaran yang dihasilkan adalah tingkat akurasi, presisi, recall, F1-Score, dan ROC-AUC dari klasifikasi serangan siber DDoS UDP Flood berbasis anomali lalu lintas jaringan.