

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Penggunaan perangkat lunak *firewall* maupun anti virus dapat membantu menahan serangan terhadap suatu host. Namun menahan serangan saja tidak cukup, terlebih bila host yang diserang merupakan suatu sistem atau server yang penting. Mengingat pentingnya masalah suatu serangan jaringan komputer, ada berbagai penelitian yang mencoba mengukur keadaan tersebut. Berdasarkan penelitian sebelumnya Nurhadi, (2017) meneliti tentang bagaimana mengklasifikasi deteksi serangan dalam jaringan dengan dataset KDD CUP 99 menggunakan metode *K-Nearest Neighbor* (K-NN) yang memperoleh hasil dengan tingkat akurasi pengujian mencapai  $k3 = 80\%$  dan  $k13 = 85,0\%$ . Namun, didalam proses teknik klasifikasi dengan menggunakan metode K-NN terdapat kekurangan dan kelemahan. Menurut Mutrofin dkk, (2014) kekurangan K-NN adalah nilai  $k$  bias, komputasi kompleks, keterbatasan memori dan mudah tertipu dengan atribut/fitur yang tidak relevan.

Maka dari itu dilakukan perbaikan dalam hal kekurangan pada metode K-NN. Salah satu perbaikan metode K-NN adalah *Modified K-Nearest Neighbor* (MK-NN). MK-NN merupakan algoritma yang dikembangkan dari algoritma K-NN, algoritma MK-NN menambahkan proses baru untuk melakukan klasifikasi yaitu perhitungan nilai validitas untuk mempertimbangkan validitas antar data latih dan perhitungan *weighted voting* untuk menghitung bobot dari masing-masing terdekat. Penambahan dua proses baru dalam MK-NN sehingga dapat memperbaiki setiap kesalahan pada proses K-NN. Penelitian Parvin dkk, (2008) menunjukkan MK-NN sangat baik dalam peningkatan akurasi dibandingkan dengan metode K-NN. Kemudian pada penelitian Maihendra, (2016) dalam memprediksi putusan perkara perceraian menggunakan metode MK-NN, didapati hasil akurasi sebesar 95,089% pada perbandingan data latih dan data uji 80:20 dengan nilai  $k=3$ . Penelitian lainnya Rasepta, (2015) mengklasifikasi status gizi

balita menggunakan metode MK-NN dengan didapati tingkat akurasi tertinggi sebesar 90% pada skenario 90:10 dan 80:20 dengan  $k = 1$  sampai  $k = 3$

Pada umumnya algoritma klasifikasi menggunakan semua fitur yang terdapat pada data untuk membangun sebuah model, padahal tidak semua fitur tersebut relevan terhadap hasil klasifikasi. Apabila hal tersebut terjadi pada data yang memiliki ukuran dan dimensi yang sangat besar, maka membuat kinerja algoritma menjadi tidak efektif dan efisien, misalnya saja waktu pemrosesan menjadi lebih lama akibat banyak fitur yang harus diproses pada dataset KDD CUP 99 yang memiliki sejumlah besar data lalu lintas jaringan yang terpantau. (Chou, Te-Shun dkk, 2007).

Salah satu solusi yang digunakan untuk mengatasi masalah tersebut adalah dengan menggunakan seleksi fitur. Seleksi fitur adalah salah satu tahap praproses pada klasifikasi. Seleksi fitur dilakukan dengan cara memilih fitur-fitur yang relevan terhadap data yang mempengaruhi hasil klasifikasi. Seleksi fitur digunakan untuk mengurangi dimensi data dan fitur yang tidak relevan, serta untuk meningkatkan efektifitas dan efisiensi kinerja dari algoritma klasifikasi. Menurut Firqiani dkk, (2007) perbandingan hasil akurasi klasifikasi data dengan seleksi fitur jauh lebih baik dari pada tanpa seleksi fitur. Rata-rata hasil akurasi data tanpa seleksi fitur yaitu 81.66% sedangkan menggunakan seleksi fitur yaitu 85.51%.

Algoritma *Fast Correlation Based Filter* adalah salah satu algoritma seleksi fitur yang dikembangkan oleh (Yu dan Huan, 2003) yang dikutip oleh (Firqiani dkk, 2005) diperoleh hasil bahwa *Fast Correlation Based Filter* sangat efisien dalam melakukan seleksi fitur serta memberikan performa yang baik bagi kinerja algoritma klasifikasi, baik dari segi waktu maupun akurasi hasil klasifikasi. Kemudian pada penelitian Chou, Te-Shun dkk (2007) membandingkan pendekatan dua algoritma pemilihan fitur berbasis korelasi yakni *Correlation-based Filter Selection* (CFS) dan *Fast Correlation-Based Filter* (FCBF) di enam kumpulan data yang diambil dari UCI databases dan KDD 99 dataset untuk melatih dan menguji algoritma pembelajaran mesin (*machine learning*) *C4.5* dan *Naive Bayes*. Dari hasil penelitian tersebut menunjukkan akurasi mencapai rata-

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

2. Dilarang mengemukakan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

rata tertinggi, memiliki kinerja superior dan dapat menseleksi fitur yang paling signifikan dalam memilih fitur informatif dari sekumpulan dataset untuk meningkatkan akurasi tugas klasifikasi.

Pada penelitian ini menggunakan dataset KDD CUP 1999 sebagai acuan data latih dan data uji. Selanjutnya data akan dipraproses menggunakan seleksi fitur *Fast Correlation Based Filter* (FCBF) dan kemudian diklasifikasi menggunakan metode *Modified K-Nearest Neighbor* (MK-NN). Diharapkan pada penelitian ini metode FCBF-MKNN mampu menghasilkan dan memperoleh akurasi yang baik dalam mengklasifikasi kelas-kelas jenis serangan pada dataset KDD CUP 99 dengan akurat.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah penelitian ini adalah bagaimana menerapkan metode *Fast Correlation Based Filter* (FCBF) sebagai tahap proses seleksi fitur yang digunakan untuk meningkatkan efektifitas dan efisiensi kinerja pada metode *Modified K-Nearest Neighbor* (MK-NN) dalam mengklasifikasi kelas-kelas jenis serangan pada dataset KDD CUP 99.

## 1.3 Batasan Masalah

Agar permasalahan yang dirumuskan dapat lebih terfokus, maka pada penelitian ini dibatasi :

1. Menggunakan dataset KDD Cup 99 yang dikeluarkan oleh DARPA (*Defense Advances Research Project Agency*).
2. Klasifikasi serangan dikategorikan kedalam 5 kelas yaitu *Denial of Service* (DoS), *Remote to Local* (R2L), *User to Root* (U2R), PROBE dan Normal.
3. Fitur awal yang digunakan sesuai dengan data yang diperoleh dari dataset KDD CUP 99 sebanyak 41 fitur.

## 1.4 Tujuan Penelitian

Tujuan pada penelitian ini adalah mengklasifikasi kelas-kelas jenis serangan pada dataset KDD CUP 99 dengan menggunakan kinerja seleksi fitur *Fast Correlation Based Filter* (FCBF) pada metode *Modified K-Nearest Neighbor* (MK-NN).

## 1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini terdiri dari pokok-pokok permasalahan yang dibahas pada masing-masing yang diuraikan menjadi beberapa bagian :

### 1.5.1 Bab I Pendahuluan

Ini membahas tentang bagaimana mendeteksi serangan dalam jaringan jaringan komputer dengan *dataset KDD CUP99* menggunakan teknik seleksi fitur algoritma *Fast Correlation Based Filter* (FCBF) pada klasifikasi metode *Modified K-Nearest Neighbor* (MK-NN) yang diangkat pada tugas akhir ini. Pada bab ini meliputi latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan sistematika penulisan.

### 1.5.2 Bab II Landasan Teori

Bab ini berisikan bagian yang menjadi landasan teori dan mendukung penulisan tugas akhir ini sebagai dasar untuk pemecahan masalah.

### 1.5.3 Bab III Metodologi Penelitian

Bab ini membahas langkah-langkah yang dilaksanakan dalam proses penelitian yaitu melakukan pengumpulan data, analisa kebutuhan sistem, perancangan perangkat lunak, implementasi dan pengujian sistem.

### 1.5.4 Bab IV Analisa dan Perancangan

Bab ini berisi pembahasan mengenai kebutuhan analisa dan perancangan untuk mengklasifikasi kelas-kelas jenis serangan pada dataset KDD CUP 99 menggunakan proses seleksi fitur *Fast Correlation Based Filter* (FCBF) dan metode klasifikasi *Modified K-nearest Neighbor* (MK-NN).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## 1.5.5 Bab V Implementasi dan Pengujian

Bab ini berisi penjelasan mengenai implementasi yang terdiri dari : batasan implementasi, lingkungan implementasi, hasil implementasi, pengujian analisa dan kesimpulan pengujian.

## 1.5.6 Bab VI Penutup

Bagian ini berisi kesimpulan yang dihasilkan dari pembahasan penelitian tentang penerapan seleksi fitur *Fast Correlation Based Filter* (FCBF) pada metode klasifikasi *Modified K-Nearest Neighbor* (MK-NN) untuk mengklasifikasi kelas jenis serangan pada dataset KDD CUP 99.