

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi pada saat ini yang memiliki peran penting bagi manusia dan berkembang cepat dari waktu ke waktu dapat memiliki resiko tersendiri, setiap orang mempunyai ketergantungan terhadap teknologi dan memiliki data yang ada pada jaringan seperti internet dalam bidang-bidang kehidupan seperti pendidikan, bisnis, dan lain-lain (Belavagi and Munial, 2016). Jaringan internet merupakan sebuah bagian dari perkembangan teknologi yang saat ini umum digunakan untuk mencari informasi secara cepat. Pertumbuhan luar biasa dan penggunaan jaringan internet yang semakin banyak menimbulkan kekhawatiran terhadap bagaimana cara berkomunikasi dan melindungi informasi digital dengan aman (Tiwari, 2017). Penelitian dalam kasus keamanan jaringan semakin banyak dilakukan dikarenakan kekhawatiran terhadap masalah yang ada akan semakin tinggi, untuk itu diperlukan sebuah sistem keamanan jaringan yang baik.

Intrusion Detection System (IDS) adalah sebuah sistem penting pada keamanan jaringan yang dapat memonitor serta dapat menemukan jika ada aktivitas yang berbahaya dan mampu mencegah intrusi dari penyusup (Tiwari, 2017). IDS berusaha menentukan apakah aktivitas pengguna yang di awasi dan lalu lintas jaringan berbahaya atau tidak. Jika serangan bahaya terdeteksi maka akan timbul sebuah peringatan. Terdapat beberapa bentuk teknik yang tersedia pada IDS untuk membedakan serangan berbahaya seperti deteksi anomali atau tanda-tanda terhadap sebuah serangan. Salah satu faktor utama yang menjadikan IDS berhasil digunakan adalah berdasarkan kualitas *feature selection* yang digunakan. IDS hanya dapat dibangun jika terdapat ketersediaan *dataset* yang efektif (Tiwari, 2017).

Network Security Layer-Knowledge Discovery in Database (NSL-KDD) adalah data terbaik versi kelanjutan dari *dataset KDDcup⁹⁹set* dan menjadi titik acuan untuk menguji sistem kerja IDS yang memiliki beberapa keunggulan dari versi sebelumnya dan bertujuan untuk mempelajari apakah algoritma klasifikasi dalam mendeteksi anomali pada pola lalu lintas jaringan efektif atau tidak (Dhanabal and Shantharajah, 2015). Penelitian yang menggunakan NSL-KDD

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.



Hak Cipta Diindungi 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.

dataset seperti yang dilakukan oleh (Dhanabal and Shantharajah, 2015), penelitian tersebut menggunakan 2 buah metode klasifikasi yaitu *Support Vector Machine* (SVM) dan *Naive Bayes*. Penelitian tersebut menggunakan *dataset* NSL-KDD karena data itu merupakan set data kandidat terbaik untuk menguji sistem kerja IDS. Akan tetapi masih terdapat kekurangan di dalam *dataset* tersebut yaitu banyaknya fitur yang berlebihan dan tidak memberikan pengaruh kepada label kelas sehingga diterapkan algoritma *feature selection* berbasis *wrapper* yaitu *Correlation-based Feature Selection* (CFS) untuk meningkatkan nilai akurasi dan mengurangi waktu pendeteksian. Berdasarkan penelitian tersebut dilihat bahwa walaupun *dataset* ini banyak sekali digunakan dalam beberapa penelitian, akan tetapi ada beberapa fitur yang terdapat pada *dataset* ini tidak semuanya memberikan pengaruh kepada label kelas, di dalam *dataset* ini terdapat beberapa fitur yang berlebihan sehingga pada saat mendeteksi intrusi memakan waktu yang lama, untuk itu diperlukan algoritma tambahan pendukung kinerja yaitu *feature Selection*.

Feature Selection merupakan teknik *pre-processing* dalam penambahan data dan pemilihan atribut yang berguna untuk menghapus fitur-fitur yang tidak relevan dan berlebihan dari NSL-KDD *dataset* (Seth, 2017). Penelitian yang menerapkan algoritma *feature selection* salah satunya dilakukan oleh (Garg and Kumar, 2014). Penelitian tersebut menggabungkan *feature selection* berbasis *filter*, kombinasi *feature selection* yang digunakan adalah *Gain Ratio* dan *Symmetrical Uncertainty* kemudian selanjutnya dilakukan klasifikasi menggunakan metode *Random Committee*, *Random Forest*, *Rotation Forest*, *Random Tree*, *Random Sub Space*, *Naive Bayesian Tree (NBTree)*, *Instance Base Learning with K parameter (IBK)*, *Part* dan *Jrip*. Hasil dari penelitian ini didapati bahwa penggabungan *Symmetrical Uncertainty* dan *Gain Ratio* mendapatkan nilai akurasi yang baik sebesar 98,05% dengan metode *IBK*. Penelitian tersebut menunjukkan bahwa dengan menerapkan *feature selection* pada beberapa metode klasifikasi yang digunakan dapat meningkatkan nilai akurasi yang tinggi.

Salah satu metode klasifikasi yang banyak digunakan di dalam beberapa penelitian yaitu metode *Modified K-Nearest Neighbor* (MK-NN). Metode ini merupakan pengembangan dari metode klasifikasi *K-Nearest Neighbor* (Parvin, Alizadeh, and Minati, 2010). Penggunaan metode MK-NN salah satunya dilakukan

Hak Cipta Diindungi Undang-Undang

1. Diarangi 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.

oleh (Parvin, Alizadeh, and Minati, 2010). Pada penelitian tersebut menggunakan 5 data sampel yang berbeda dan membagi data menjadi 80% data latih dan 20% data uji untuk percobaan yang dilakukan. Hasil dari penelitian tersebut menunjukkan bahwa metode MK-NN dapat meningkatkan kinerja yang lebih baik daripada metode sebelumnya yaitu *K-Nearest Neighbor* (K-NN), karena pada metode ini melakukan penambahan nilai validitas sebagai tahap *pre-processing* untuk mendapatkan informasi tentang data latih, nilai validitas memperhitungkan nilai stabilitas dari setiap data latih dengan tetangga-tetangganya lalu menerapkan *weighted voting* untuk memproses nilai validitas, dari 5 data menggunakan $k=3, 5$ dan 7 menunjukkan peningkatan tingkat akurasi pada semua percobaan yang dilakukan.

Berdasarkan penjelasan yang telah diuraikan di atas maka akan dilakukan penelitian dengan menerapkan algoritma *feature selection* untuk klasifikasi pada metode klasifikasi *Modified K-Nearest Neighbor* (MK-NN) dengan menggunakan NSL-KDD *dataset*, metode MK-NN digunakan karena metode tersebut lebih meningkatkan kinerja daripada metode sebelumnya yaitu *K-Nearest Neighbor* (K-NN). *Feature selection* yang digunakan berbasis *filter* yaitu *Symmetrical Uncertainty* dan *Gain Ratio*. Dengan menerapkan *feature selection* pada metode yang digunakan diharapkan dapat memperoleh peningkatan nilai akurasi yang baik.

1.2 Rumusan Masalah

Penelitian yang akan dilakukan berdasarkan latar belakang yang telah diuraikan adalah:

1. Bagaimana mengimplementasikan algoritma *Symmetrical Uncertainty* dan *Gain Ratio* untuk klasifikasi serangan jaringan menggunakan metode *Modified K-Nearest Neighbor* menggunakan NSL-KDD *dataset*.
2. Bagaimana membangun sebuah sistem yang dapat melakukan klasifikasi terhadap serangan jaringan menggunakan metode klasifikasi *Modified K-Nearest Neighbor* (MK-NN).
3. Bagaimana tingkat akurasi pengklasifikasian serangan jaringan menggunakan metode klasifikasi *Modified K-Nearest Neighbor* (MK-NN).



1.3 Batasan Masalah

Terdapat ruang lingkup pada penelitian ini yang bertujuan untuk memfokuskan permasalahan yang dirumuskan, maka diberikan beberapa batasan masalah sebagai berikut:

1. Jumlah data pada penelitian ini terdapat 2 jenis data, yaitu 25 data (setara) dan 1052 (tidak setara).
2. Kelas yang digunakan dalam klasifikasi terbagi 5 kelas, yaitu *Normal*, *Denial of Service (DoS)*, *ProbeS*, *User to Root (U2R)* dan *Remote to Local (R2L)*.
3. Nilai parameter K yang digunakan adalah bilangan ganjil 1 hingga 11.
4. Perbandingan data latih dan data uji yang diterapkan adalah 70%:30%, 80%:20% dan 90%:10%.
5. Penelitian yang dilakukan tidak termasuk implementasi pada suatu sistem jaringan.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai pada penelitian yang akan dilakukan adalah:

1. Untuk menerapkan algoritma *Symmetrical Uncertainty (SU)* dan *Gain Ratio (GR)* untuk klasifikasi serangan jaringan menggunakan metode *Modified K-Nearest Neighbor* menggunakan dataset NSL-KDD.
2. Untuk menghasilkan sebuah sistem yang dapat melakukan klasifikasi terhadap serangan jaringan menggunakan metode klasifikasi *Modified K-Nearest Neighbor (MK-NN)* pada dataset NSL-KDD.
3. Untuk mengetahui hasil nilai akurasi metode *Modified K-Nearest Neighbor (MK-NN)* dengan menggunakan *feature selection Symmetrical Uncertainty (SU)* dan *Gain Ratio (GR)*.

Hak Cipta Diindungi 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.

1.5 Sistematika Penulisan

BAB I PENDAHULUAN

Bab ini berisikan tentang latar belakang penelitian, beberapa rumusan masalah, batasan-batasan masalah yang ditetapkan, tujuan dari penelitian dan sistematika penulisan laporan berdasarkan penelitian yang dilakukan.

BAB II LANDASAN TEORI

Bab ini berisikan tentang beberapa teori yang berkaitan dengan penelitian seperti dataset yang digunakan, *feature selection* yang digunakan dan metode yang digunakan yaitu MK-NN.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan beberapa langkah dan tahap yang dilakukan dalam penelitian ini. Beberapa langkah tersebut meliputi alur penelitian, pengumpulan data yang diperlukan, analisa dan perancangan sistem, hingga pada tahap implementasi dan pengujian.

BAB IV ANALISA DAN PERANCANGAN

Bab ini berisi tentang beberapa analisa yang dilakukan untuk kebutuhan sistem yang akan dibangun, serta terhadap metode yang digunakan dalam penelitian ini yaitu *Modified K-Nearest Neighbor* (MK-NN).

BAB V IMPLEMENTASI DAN PENGUJIAN

Bab ini menjelaskan tentang implementasi *feature selection* untuk mendukung metode klasifikasi yang digunakan serta hasil dari pengujian sistem.

BAB VI KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari penelitian yang dilakukan serta beberapa saran untuk pengembangan terhadap pengembangan sistem selanjutnya.

Hak Cipta Diindungi Undang-Undang

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