

**ANALISA KOMBINASI HASIL SELEKSI FITUR
SYMMETRICAL UNCERTAINTY DAN GAIN RATIO
PADA DATASET NSL-KDD MENGGUNAKAN METODE
*NAÏVE BAYES***

FAUZUL AZMI

11251104826

Jurusan Teknik Informatika

Fakultas Sians dan Teknologi

Universitas Islam Negeri Sultan Syarif Kasim Riau

ABSTRAK

Intrusion Detection System (IDS), merupakan sistem atau perangkat yang dapat melakukan deteksi terhadap serangan dalam jaringan komputer. Sebagai langkah awal membangun IDS, hampir 70% penelitian dilakukan menggunakan dataset *Network Security Layer - Knowledge Discovery in Database* (NSL-KDD). Walaupun demikian, dataset NSL-KDD memiliki fitur yang tidak relevan dan berlebihan, menyebabkan waktu proses yang lama dan hasil *detection rate* rendah. Untuk mengatasi permasalahan tersebut, maka diterapkanlah algoritma *Feature selection*, yaitu algoritma untuk mengurangi fitur atau atribut yang ada pada dataset. *Feature selection* yang akan digunakan pada penelitian ini adalah *Gain Ratio* (GR) dan *Symmetrical Uncertainty* (SU), hasil *feature selection* kemudian dikombinasikan menggunakan persamaan *Boolean And*. Adapun tujuan dari penelitian ini adalah menganalisa kombinasi hasil *Feature selection* pada dataset NSL-KDD menggunakan model klasifikasi *Naïve Bayes*. Dataset NSL-KDD yang akan digunakan sebelumnya akan disetarakan jumlah pada setiap kelasnya, kemudian didiskritisasi, kemudian diseleksi atributnya menggunakan Algortima *Feature selection* GR dan SU, selanjutnya 20 atribut dengan nilai terbaik dari hasil masing-masing algoritma *Feature selection* dikombinasikan menggunakan persamaan *Boolean AND*, sehingga terpilihlah 16 atribut dari kumpulan atribut yang ada. Setelah dianalisa menggunakan model klasifikasi *Naïve Bayes*, dapat disimpulkan bahwa kombinasi hasil *feature selection* SU dan GR pada dataset NSL-KDD dapat memberikan hasil yang baik dengan tingkat akurasi rata-rata mencapai 90%.

Kata kunci : *Boolean AND, Feature selection, Gain Ratio, Intrusion Detection System (IDS), Naïve bayes, NSL-KDD, Symmetrical Uncertainty.*

***COMBINATION ANALYSIS OF SELECTION OF
SYETETRIS UNCERTAINTY AND GAIN RATIO
FEATURES IN NSL-KDD DATASET USING NAÏVE
BAYES METHOD***

FAUZUL AZMI

11251104826

Informatic Engineering Department

Faculty of Science and Technology

State Islamic University of Sultan Syarif Kasim Riau

ABSTRACT

Intrusion Detection System (IDS), is a system or device that can detect attacks in computer networks. As a first step in building IDS, almost 70% of the research was conducted using datasets Network Security Layer - Knowledge Discovery in Database (NSL-KDD). However, the NSL-KDD dataset has features that are irrelevant and excessive, causing long processing times and low detection rate results. To overcome these problems, the Feature selection algorithm is applied, an algorithm to reduce the features or attributes that exist in the dataset. The Feature selection that will be used in this study is Gain Ratio (GR) and Symmetrical Uncertainty (SU), the results of the feature selection are then combined using the Boolean equation. The purpose of this study is to analyze the combination of Feature selection results in the NSL-KDD dataset using the classification model Naïve Bayes. The NSL-KDD dataset that will be used before will be equalized in each class, then discretized, then selected attributes using GR and SU Feature selection algorithms, then 20 attributes with the best value from the results of each Feature selection algorithm combined using AND Boolean equations, so 16 attributes are selected from the existing set of attributes. After being analyzed using the Naïve Bayes classification model, it can be concluded that the combination of the results of SU and GR feature selection in the NSL-KDD dataset can provide good results with an average accuracy rate of 90%.

Keyword : Boolean AND, Feature selection, Gain Ratio, Intrusion Detection System (IDS), Naïve bayes, NSL-KDD, Symmetrical Uncertainty.