REDUKSI DIMENSI DATA MENGGUNAKAN KOMBINASI METODE SELEKSI FITUR FILTER VARIAN RENDAH DAN *RELIEFF* DALAM PREDIKSI SERANGAN *BOTNET*

TESIS

OLEH

NUR DWI MURYANTO

NIM: 55419110012

PROGRAM STUDI MAGISTER TEKNIK ELEKTRO

FAKULTAS TEKNIK

UNIVERSITAS MERCU BUANA

2021

REDUKSI DIMENSI DATA MENGGUNAKAN KOMBINASI METODE SELEKSI FITUR FILTER VARIAN RENDAH DAN *RELIEFF* DALAM PREDIKSI SERANGAN *BOTNET*

TESIS

Diajukan sebagai Salah Satu Syarat untuk Menyelesaikan
Program Studi Magister Teknik Elektro

OLEH
NUR DWI MURYANTO
NIM: 55419110012

PROGRAM STUDI MAGISTER TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS MERCU BUANA
2021

ABSTRAK

Botnet merupakan salah satu jenis *malware* yang menjadi ancaman serius dalam dunia keamanan siber, terutama dalam perangkat *IoT.* Oleh karena itu, diperlukan upaya prediksi terhadap suatu serangan *Botnet.* Salah satu teknik yang digunakan adalah *Machine learning-based*, yang berbasis data (pengetahuan) untuk dipelajari dan dibuat suatu model matematis. Namun, permasalahan umum dalam prediksi serangan di era *big data* adalah dimensi *Dataset* yang besar. Hal ini berpengaruh pada kecepatan komputasinya, sementara itu ada kemungkinan tidak semua fitur relevan dalam prediksi serangan. Oleh karena itu, dalam penelitian ini akan dilakukan reduksi fitur dalam *Dataset* menggunakan kombinasi metode seleksi fitur Filter Varian Rendah dan *ReliefF* pada algoritma *machine learning Support Vector Machine*. Saat Filter Varian Rendah dilakukan di depan *ReliefF* (Skenario 2), terjadi peningkatan kecepatan komputasi dengan nilai rata-rata mencapai 39,1 kali lebih cepat dibandingkan tanpa menggunakan seleksi fitur (Skenario 1). Sedangkan jika diterapkan *ReliefF* sebelum Filter Varian Rendah (Skenario 3) mencapai rata-rata 30,91 kali lebih cepat. Selain itu, evaluasi model menunjukkan peningkatan Akurasi, Presisi, dan *Specificity*, yaitu peningkatan Akurasi, dengan rata-rata 38,627% (Skenario 2) dan 38,059% (Skenario 3); peningkatan Presisi, dengan rata-rata 42,839% (Skenario 3); dan 42,452% (Skenario 3), serta peningkatan *Specificity* dengan rata-rata 81,14% (Skenario 2) dan 84,336% (Skenario 3). Namun terjadi penurunan *Recall* meskipun tidak signifikan dengan rata-rata sebesar 4,453% (Skenario 2) dan 5,844% (Skenario 3).

Kata Kunci: *Botnet*, SVM, Seleksi Fitur, Filter Varian Rendah, *ReliefF*

.

ABSTRACT

Botnet is one kind of malware that become serious threat in cyber security, especially in IoT devices. Therefore, it is necessary to predict a botnet attack. One of the techniques used is Machine learning-based, which is based on data (knowledge) to be studied and made a mathematical model. However, a common problem in predicting attacks in the era of big data is the large dimension of the Dataset which affects its computational speed, while it is possible that not all features are relevant in attack prediction. Therefore, in this study proposed feature reduction in the Dataset will using a combination of the Low Variant Filter feature selection method and ReliefF on the machine learning algorithm Support Vector Machine. Therefore, in this study, a feature reduction method in the Dataset is proposed using a combination of two feature selection methods, namely Low Variance Filter and ReliefF, on the Support Vector Machine. When the Low Variance Filter was performed in front of ReliefF (Scenario 2), there was an increase in computational speed with an average value of 39.1 times faster than without using feature selection (Scenario 1). Meanwhile, if ReliefF is applied before the Low Variant Filter (Scenario 3), it reaches an average of 30.91 times faster. In addition, the model evaluation shows an increase in Accuracy, Precision, and Specificity, namely an increase in accuracy, with an average of 38.627% (Scenario 2) and 38.059% (Scenario 3); increased precision, with an average of 42.839% (Scenario 3); and 42.452% (Scenario 3), as well as an increase in Specificity with an average of 81.14% (Scenario 2) and 84.336% (Scenario 3). However, there was a decrease in recall although it was not significant with an average of 4.453% (Scenario 2) and 5.844% (Scenario 3).

Keywords: *Botnet, SVM, Feature Selection, Low Variance Filter, ReliefF*

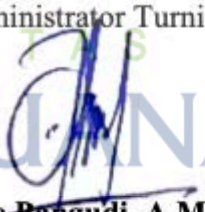
PERNYATAAN *SIMILARITY CHECK*

Saya yang bertanda tangan di bawah ini menyatakan, bahwa karya ilmiah yang ditulis oleh

Nama	: Nur Dwi Muryanto
NIM	: 55419110012
Program Studi	: Magister Teknik Elektro

dengan judul
"*Reduksi Dimensi Data Menggunakan Kombinasi Metode Seleksi Fitur Filter Varian Rendah dan ReliefF dalam Prediksi Serangan Botnet*", telah dilakukan pengecekan similarity dengan sistem Turnitin pada tanggal 05/08/2021, didapatkan nilai persentase sebesar 18 %.

Jakarta, 05 Agustus 2021
Administrator Turnitin



Arie Pangudi, A.Md

PENGESAHAN

Judul	:	Reduksi Dimensi Data menggunakan Kombinasi Metode Seleksi Fitur Filter Varian Rendah dan *ReliefF* dalam Prediksi Serangan *Botnet*
Bentuk Tesis	:	Penelitian Kuantitatif
Nama	:	Nur Dwi Muryanto
NIM	:	55419110012
Program	:	Magister Teknik Elektro
Tanggal	:	30 Agustus 2021

Mengesahkan

Pembimbing

Dr. Marza Ihsan Marzuki, MT.

Dekan Fakultas Teknik	Ketua Program Studi Magister Teknik Elektro
Dr. Ir. Mawardi Amin, M.T.	Dr. Umaisaroh, S.ST

PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan dengan sebenar-benarnya bahwa semua pernyataan dalam Tesis ini :

Judul	:	Reduksi Dimensi Data Menggunakan Kombinasi Metode Seleksi Fitur Filter Varian Rendah dan *ReliefF* dalam Prediksi Serangan *Botnet*
Bentuk Tesis	:	Penelitian Kuantitatif
Nama	:	Nur Dwi Muryanto
NIM	:	55419110012
Program	:	Magister Teknik Elektro
Tanggal	:	30 Agustus 2021

Merupakan hasil studi pustaka, penelitian lapangan, dan karya saya sendiri dengan bimbingan Dosen Pembimbing yang ditetapkan dengan Surat Keputusan Program Studi Magister Teknik Elektro Universitas Mercu Buana.

Karya ilmiah ini belum pernah diajukan untuk memperoleh gelar kesarjanaan pada program sejenis di perguruan tinggi lain. Semua informasi, data, dan hasil pengolahannya yang digunakan, telah dinyatakan secara jelas sumbernya dan dapat diperiksa kebenarannya.

Jakarta, 30 Agustus 2021

Nur Dwi Muryanto

KATA PENGANTAR

Segala puji dan syukur senantiasa dipanjatkan atas kehadirat Allah SWT yang telah melimpahkan rahmat, nikmat, dan karunia-Nya kepada kita. Shalawat dan salam semoga senantiasa tercurah kepada suri tauladan kita, Rasulullah Muhammad SAW. Syukur Alhamdulillah penulis ucapkan atas terselesaikannya penulisan Tesis dengan judul "**Reduksi Dimensi Data Menggunakan Kombinasi Metode Seleksi Fitur Filter Varian Rendah dan *ReliefF* dalam Prediksi Serangan *Botnet***" sebagai syarat kelulusan dan mendapatkan gelar Magister Teknik Elektro pada Program Pascasarjana di Universitas Mercu Buana.

Penulis menyadari masih banyaknya kekurangan dan keterbatasan dalam proses penelitian ini. Selama proses penelitian, penulis banyak menerima bimbingan, dukungan, dan bantuan baik materiil dan immateriil dari berbagai pihak, untuk itu ijinkan penulis untuk mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Istri tercinta dan keluarga yang memberikan semangat dan dukungannya kepada penulis dalam penyelesaian penelitian Tesis ini;
2. Marza Ihsan Marzuki, MT., Ph.D, selaku pembimbing penulisan Tesis ini yang telah mencurahkan waktunya untuk memberikan arahan, bimbingan, dan masukan yang membangun dalam penyelesaian penulisan Tesis ini;
3. Prof. Dr. Ir. Andi Adriansyah, M.Eng dan Dr. Umaisaroh, S.ST selaku Kepala Program Studi Magister Teknik Elektro yang selalu mengingatkan kami untuk semangat dalam penyelesaian Tesis ini;
4. Segenap dosen pengajar Program Pascasarjana Magister Teknik Elektro Universitas Mercu Buana yang telah membagikan ilmu, pengetahuan, wawasan, dan pengalamannya yang sangat bermanfaat bagi pengembangan diri kami;
5. Segenap staff Tata Usaha, khususnya Program Pascasarjana Magister Teknik Elektro, yang telah membantu kami dalam segala proses administrasi pendidikan;

6. Segenap pimpinan di BSSN yang telah mengizinkan penulis untuk melanjutkan pendidikan ke jenjang Magister;
7. Rekan-rekan seperjuangan di MTEL-25 yang telah saling mendukung demi terselesaikannya Tesis ini.

Penulis menyadari bahwa dalam penulisan Tesis ini masih terdapat banyak kekurangan. Oleh karena itu, Penulis menerima saran dan kritik yang membangun. Penulis berharap bahwa penulisan Tesis ini dapat memberikan manfaat yang sebesar-besarnya bagi kemajuan ilmu pengetahuan di Indonesia, khususnya dalam bidang Keamanan Siber.

Jakarta, 30 Agustus 2021

Nur Dwi Muryanto

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR ISTILAH

Akurasi (*Accuracy*) Suatu nilai yang merepresentasikan prosentase dari keseluruhan data yang diprediksi dengan benar (TP dan TN) dari keseluruhan data uji.

Akurasi dirumuskan sebagai berikut

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Binary Classification Klasifikasi suatu himpunan input ke dalam satu dari dua buah kelas. Misal klasifikasi trafik jaringan menjadi trafik normal atau serangan.

Botnet Salah satu jenis *malware* yang berasal dari kata "*robot*" dan "*network*". *Botnet* akan membuat suatu infrastruktur komunikasi privat yang dapat digunakan untuk tujuan jahat, yaitu untuk mengontrol sejumlah besar komputer yang telah terpasang *backdoor* di dalamnya.

Confusion Matrix Sebuah tabel berukuran NxN yang berisi rangkuman hasil prediksi dari suatu model klasifikasi, baik prediksi yang benar maupun yang salah, untuk melihat seberapa sukses model dalam melakukan klasifikasi.

Data Cleaning Tahapan pada Pra-proses yang digunakan untuk membuang data noise, seperti data bernilai kosong (null), data tidak konsisten, dan data berulang.

Data Latih Bagian dari *Dataset* yang digunakan untuk melatih model.

Data Uji Bagian dari *Dataset* yang digunakan untuk menguji model yang dihasilkan dari proses pelatihan data.

Dataset Sekumpulan data yang digunakan dalam proses pelatihan dan pengujian data dalam *machine learning*.

False Negative (FN) Hasil prediksi model yang salah memprediksi kelas positif sebagai kelas negatif. Misal: suatu trafik serangan yang diprediksi sebagai trafik normal oleh model.

False Positive (FP) Hasil prediksi model yang salah memprediksi kelas negatif sebagai kelas positif. Misal: suatu trafik normal yang diprediksi sebagai trafik serangan.

Fitur Kolom dalam *Dataset* yang digunakan sebagai variabel masukan yang digunakan dalam membuat pemodelan prediksi. Memiliki nama lain atribut. Sebagian besar fitur digunakan sebagai inputan dan sebagian (umumnya 1 fitur/kolom) digunakan sebagai output atau label prediksi.

Instance Prinsipnya, setiap baris *Dataset* disebut instance yang menunjukkan pengamatan dari domain permasalahan.

Machine learning Bagian dari keilmuan *Artificial Intelligent* (kecerdasan buatan) yang erat kaitannya dengan komputer yang menggunakan teknik statistika untuk mempelajari sekumpulan data sehingga komputer belajar sendiri, berfikir sendiri, dan mengambil keputusan sendiri.

Malware Suatu jenis perangkat lunak (software) yang digunakan untuk tujuan jahat (*Malicious Software*).

Model Hasil pemodelan matematis sebagai representasi dari apa yang telah dipelajari oleh sistem dari data latih.

Multi-Class Clasification klasifikasi suatu himpunan input ke dalam salah satu dari banyak kelas (lebih dari dua kelas) yang ditentukan. Misal: terdapat beberapa jenis trafik, yaitu normal, *DoS, DDos*, dan *Theft*, maka *Dataset* akan diklasifikasikan dalam salah satu dari keempat kelas tersebut oleh model.

One-Hot Encoding Salah satu metode penanganan data kategorik yang memetakan elemen unik dalam kolom *Dataset* menjadi sebuah variabel dummy, jika pada baris data terdapat elemen tersebut maka nilainya 1, jika bukan maka nilainya 0.

Overfitting Suatu model yang sangat akurat pada data latih, namun saat menggunakan data baru, model gagal dalam membuat prediksi secara benar.

Presisi (*Precition*) Suatu nilai yang merepresentasikan mengidentifikasi frekuensi di mana model itu benar saat memprediksi kelas positif.

Presisi dirumuskan sebagai berikut

$$Precision = \frac{TP}{TP + FP}$$

Pra-Proses Proses awal sebelum *Dataset* diolah untuk dilakukan proses pelatihan dan pengujian data.

Recall Suatu nilai yang merepresentasikan prosentase dari label positif yang diprediksi seara benar dari seluruh label positif pada data uji.

Recall dirumuskan sebagai berikut

$$Recall = \frac{TP}{TP + FN}$$

Skenario 1 Skenario penelitian tanpa menggunakan metode seleksi fitur.

Skenario 2 Skenario penelitian dengan menggunakan kombinasi seleksi fitur, dimana dilakukan Filter Varian Rendah dilanjutkan *ReliefF*.

Skenario 3 Skenario penelitian dengan menggunakan kombinasi seleksi fitur, dimana dilakukan *ReliefF* dilanjutkan Filter Varian Rendah.

Skenario 4 Skenario penelitian dengan menggunakan metode seleksi fitur Filter Varian Rendah.

Skenario 5 Skenario penelitian dengan menggunakan metode seleksi fitur *ReliefF*.

Supervised Learning Suatu teknik dalam melatih model dari data masukan yang menyertakan label yang sesuai.

Specificity Nilai yang merepresentasikan prosentase model dalam memprediksi data negatif dengan benar dari keseluruhan data negatif pada data uji.

Specificity dirumuskan sebagai berikut

$$Specificity = \frac{TN}{FP + TN}$$

True negative (TN) Hasil prediksi model yang dengan benar memprediksi kelas negatif. Misal: trafik normal diprediksi dengan benar sebagai trafik normal.

True positive (TP) Hasil prediksi model yang dengan benar memprediksi kelas positif. Misal: trafik serangan diprediksi dengan benar sebagai trafik serangan.