



**ANALISIS SELEKSI FITUR DALAM MENENTUKAN
SERANGAN MENGGUNAKAN ALGORITMA *DECISION*
*TREE***

SKRIPSI

**HIDAYAH HUSNUL KHOTIMAH
1610511057**

UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2020



**ANALISIS SELEKSI FITUR DALAM MENENTUKAN
SERANGAN MENGGUNAKAN ALGORITMA *DECISION*
*TREE***

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
Sarjana Komputer**

HIDAYAH HUSNUL KHOTIMAH

1610511057

UNIVERSITAS PEMBANGUNAN NASIONAL “ VETERAN” JAKARTA

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI INFORMATIKA

2020

PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Hidayah Husnul Khotimah

NIM : 1610511057

Tanggal : 17 Mei 2020

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 17 Mei 2020

Yang Menyatakan,



(Hidayah Husnul K)

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Hidayah Husnul Khotimah

NIM : 1610511057

Fakultas : Ilmu Komputer

Program Studi : S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Analisis Seleksi Fitur dalam Menentukan Serangan Menggunakan Algoritma *Decision Tree*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada Tanggal : 17 Mei 2020
Yang Menyatakan,



(Hidayah Husnul K)

LEMBAR PENGESAHAN

Dengan ini dinyatakan bahwa Skripsi berikut:

Nama : Hidayah Husnul Khotimah
NIM : 1610511057
Program Studi : S1 Informatika
Judul Skripsi : Analisis Seleksi Fitur dalam Menentukan Serangan Menggunakan Algoritma *Decision Tree*

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



Yuni Widiastiwi, S.Kom., M.Si.
Penguji I



I Wayan Widi Pradnyana, S.Kom.,
MTI
Penguji II



Henki Bayu Seta, S.Kom., MTI.
Pembimbing I



Nurul Chamidah, S.Kom., M.Kom.
Pembimbing II



Ermatita, M.Kom.
Dekan



Anita Muliawati, S.Kom., MTI.
Ketua Program Studi

Ditetapkan di : Jakarta
Tanggal Persetujuan : 17 Juni 2020



ANALISIS SELEKSI FITUR DALAM MENENTUKAN SERANGAN MENGGUNAKAN ALGORITMA *DECISION* *TREE*

Hidayah Husnul Khotimah

ABSTRAK

Kejahatan siber (*cyber crime*) adalah kejahatan pada dunia maya yang memanfaatkan teknologi dan telekomunikasi sebagai medianya. Salah satu kejahatan siber adalah serangan siber (*cyber attack*). Serangan siber merupakan serangan terhadap sistem informasi, infrastruktur, jaringan komputer, serta sistem keamanan komputer. Oleh karena itu, perlu adanya alat yang dapat digunakan untuk mendeteksi serangan yang disebut sebagai *Intrusion Detection System* (IDS). Hasil dari IDS ini berupa *log* serangan yang nantinya akan dianalisa menggunakan *Security Event Management* (SEM) oleh administrator keamanan jaringan. Agar administrator keamanan jaringan lebih mudah dalam menganalisa SEM, diperlukan metode seleksi fitur yang berfungsi untuk menentukan fitur-fitur apa saja yang berpengaruh dalam mendeteksi serangan. Metode yang digunakan dalam seleksi fitur adalah *ranking importance* dengan algoritma klasifikasi *decision tree* sebagai penghitungan akurasi. Dari hasil penelitian didapat fitur-fitur penting, yang terdiri dari *source ip*, *source port*, *signature id*, *rev*, *signature*, *severity*, *source city name*, *destination regional name*, dan *destination city name* dengan akurasi klasifikasi sebesar 99.97%, waktu pelatihan selama 0.021208 s, serta waktu pengujian selama 0.001554 s.

Kata kunci: Kejahatan Siber, Serangan Siber, IDS, SEM, *ranking importance*, *Decision Tree*.

ANALYSIS OF FEATURES SELECTION IN DETERMINING THE ATTACK USING DECISION TREE ALGORITHM

Hidayah Husnul Khotimah

ABSTRACT

Cybercrime is a crime in cyberspace that utilizes technology and telecommunications as its medium. One of the cybercrime is cyber attack. Cyber attack is an attack on information systems, infrastructure, computer networks, and computer security systems. Therefore, there needs to be a tool that can be used to detect an attack called the Intrusion Detection System (IDS). The result of this IDS is in the form of an attack log which will later be analyzed using Security Event Management (SEM) by the network security administrator. To make network security administrators easier to analyze SEM, a feature selection method is needed that functions to determine what features are influential in detecting attacks. The method used in feature selection is importance ranking with the decision tree classification algorithm as an accuracy calculation. From the research results obtained important features, which consist of source ip, source port, signature id, rev, signature, severity, source city name, destination regional name, and destination city name with classification accuracy of 99.97%, training time for 0.021208 s, as well as the testing time of 0.001554 s.

Keyword: Cybercrime, Cyber Attack, IDS, SEM, ranking importance, Decision Tree.

KATA PENGANTAR

Puji dan syukur atas kehadiran Allah SWT atas segala rahmat dan karunia-Nya, shalawat serta salam tak lupa kita haturkan kepada Nabi Muhammad SAW, keluarga serta sahabatnya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “**Analisis Seleksi Fitur Dalam Menentukan Serangan Menggunakan Algoritma *Decision Tree***”.

Penulisan Tugas Akhir ini merupakan salah satu syarat untuk memperoleh gelar sarjana Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jakarta. Rasa terimakasih penulis ucapkan kepada:

1. Kepada orang tua dan keluarga yang saya cintai, yang telah memberikan dukungan dan doa kepada penulis selama masa pengerjaan skripsi ini.
2. Ibu Dr. Ermatita, M.Kom., selaku Dekan Fakultas Ilmu Komputer
3. Ibu Anita Muliawati, S.Kom., MTI., selaku Kepala Program Studi Informatika.
4. Bapak Henki Bayu Seta, S.Kom., MTI., dan Ibu Nurul Chamidah, S.Kom., M.Kom., selaku dosen pembimbing I & II skripsi yang membantu penulis dalam penyusunan skripsi sehingga dapat menyelesaikan skripsi dengan baik.
5. Ibu Ika Nurlaili Isnainiyah, S.Kom., M.Sc., selaku dosen pembimbing akademik.
6. Ibu, Bapak Dosen Program Studi Informatika UPN “Veteran” Jakarta atas ilmu-ilmu yang bermanfaat.
7. Kepada Pihak Instansi, yang telah memberikan izin untuk mengambil data *Log IDS Suricata*.
8. Kepada teman-teman penulis mahasiswa Program Studi Informatika Angkatan 2016 yang tidak dapat penulis sebutkan namanya satu persatu.

Akhir kata, semoga skripsi ini dapat bermanfaat bagi para pembacanya.

Jakarta, 18 Mei 2020

Penulis

Hidayah Husnul Khotimah

DAFTAR ISI

ANALISIS SELEKSI FITUR DALAM MENENTUKAN SERANGAN MENGUNAKAN ALGORITMA <i>DECISION TREE</i>	ii
PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
KATA PENGANTAR	i
DAFTAR ISI.....	ii
DAFTAR TABEL.....	v
DAFTAR GAMBAR	vi
DAFTAR LAMPIRAN.....	viii
DAFTAR SIMBOL.....	ix
1. BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Ruang Lingkup	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.6 Luaran yang Diharapkan	3
1.7 Sistematika Penulisan.....	3
2. BAB II.....	5
2.1 Kejahatan Siber (<i>Cyber Crime</i>).....	5
2.1.1 Definisi Kejahatan Siber (<i>Cyber Crime</i>)	5
2.1.2 Jenis-Jenis Kejahatan Siber (<i>Cyber Crime</i>).....	5
2.2 Serangan Siber (<i>Cyber Attack</i>).....	7
2.3 <i>Security Event Management (SEM)</i>	7
2.4 <i>Intrusion Detection System (IDS)</i>	8
2.4.1 Definisi <i>Intrusion Detection System (IDS)</i>	8
2.4.2 Jenis-Jenis <i>Intrusion Detection System (IDS)</i>	8

2.5	Suricata.....	9
2.6	<i>Log</i>	10
2.7	Seleksi Fitur.....	10
2.8	<i>Ranking Importance</i>	11
2.9	Klasifikasi.....	12
2.9.1	<i>Decision Tree</i>	13
2.9.2	Ukuran Pemilihan Atribut.....	14
2.10	Studi Literatur.....	16
3.	BAB III	19
3.1	Kerangka Pikir.....	19
3.1.1	Studi Pustaka.....	19
3.1.2	Identifikasi Masalah	20
3.1.3	Pengumpulan Data	20
3.1.4	Data Parsing	20
3.1.5	Pembersihan Data.....	20
3.1.6	Transformasi Data.....	20
3.1.7	Seleksi Fitur	20
3.1.8	Pembagian Data	21
3.1.9	Pemodelan Data	21
3.1.10	Pengujian Data	21
3.1.11	Evaluasi.....	22
3.2	Perangkat Penelitian	22
3.3	Tempat Penelitian.....	22
3.4	Jadwal Penelitian	22
4.	BAB IV	24
4.1	Pengumpulan Data	24
4.2	Data Parsing	24
4.3	Pembersihan Data.....	27
4.3.1	Penghapusan Baris	27
4.3.2	Penghapusan Kolom (Atribut)	28
4.3.3	<i>Missing Value</i>	30
4.4	Transformasi Data	32

4.5	Pembagian Data.....	34
4.5.1	Data Latih.....	35
4.5.2	Data Uji.....	35
4.6	Decision Tree	36
4.6.1	Pembentukan Pohon.....	36
4.6.2	Pembentukan Aturan Keputusan.....	37
4.6.3	Pengujian Metode Decision Tree	38
4.6.4	Hasil Prediksi	39
4.7	Seleksi Fitur.....	40
4.7.1	Decision Tree untuk Seleksi Fitur.....	40
4.7.2	Perbandingan dan Hasil Seleksi Fitur	42
4.8	Klasifikasi dengan Seleksi Fitur.....	44
4.9	ELK (<i>Elasticsearch Logstash Kibana</i>).....	46
4.9.1	<i>Destination Port</i> Berdasarkan <i>Timestamp</i>	46
4.9.2	<i>Severity</i> dan <i>Category</i> Berdasarkan <i>Source Country</i>	46
4.9.3	<i>Destination Port</i> dan <i>Siganture</i> Berdasarkan <i>Severity</i>	48
4.9.4	<i>Destination Country</i>	49
4.9.5	Analisis Serangan yang Terjadi	50
5.	BAB V	51
5.1	Kesimpulan.....	51
5.2	Saran.....	51
6.	DAFTAR PUSTAKA	52
	RIWAYAT HIDUP.....	54
	LAMPIRAN.....	55

DAFTAR TABEL

Tabel 3.1 Jadwal Penelitian.....	23
Tabel 4.1 Data <i>Log</i> IDS Hasil <i>Parsing</i> (CSV).....	24
Tabel 4.2 Keterangan Atribut Data <i>Log</i> IDS	25
Tabel 4.3 Atribut Data <i>Log</i> IDS Setelah Dihapus.....	29
Tabel 4.4 Huruf Kapital pada <i>Label Encoding</i>	32
Tabel 4.5 Spasi pada <i>Label Encoding</i>	33
Tabel 4.6 Nilai Minus pada <i>Label Encoding</i>	33
Tabel 4.7 <i>10-fold Cross Validation</i> Untuk Data <i>Log</i> IDS.....	35
Tabel 4.8 Hasil Klasifikasi dengan Semua Fitur.....	39
Tabel 4.9 Hasil Prediksi dengan Decision Tree	39
Tabel 4.10 Aturan Seleksi Fitur	40
Tabel 4.11 Akurasi dan Waktu Pelatihan dari Fitur yang Dihapus.....	41
Tabel 4.12 Hasil Aturan Seleksi Fitur.....	42
Tabel 4.13 Hasil Seleksi Fitur.....	43
Tabel 4.14 Hasil Klasifikasi dengan Fitur Penting	45

DAFTAR GAMBAR

Gambar 2.1 <i>Security Event Management</i>	8
Gambar 2.2 Suricata.....	9
Gambar 2.3 <i>Log IDS</i>	10
Gambar 2.4 Konsep <i>Decision Tree</i>	13
Gambar 2.5 Struktur <i>Decision Tree</i>	14
Gambar 3.1 Metodologi Penelitian	19
Gambar 3.2 <i>10-fold Cross Validation</i>	21
Gambar 4.1 Data Awal <i>Log IDS</i> (JSON).....	24
Gambar 4.2 Baris yang Bernilai Kosong pada Data <i>Log IDS</i>	27
Gambar 4.3 Atribut yang Bernilai Kosong pada Data <i>Log IDS</i>	28
Gambar 4.4 Atribut yang Bernilai Sama dalam 1 Kolom pada Data <i>Log IDS</i>	29
Gambar 4.5 Data <i>Log IDS</i> Sebelum Dilakukan <i>Reverse Geocoding</i>	31
Gambar 4.6 Data <i>Log IDS</i> Setelah Dilakukan <i>Reverse Geocoding</i>	31
Gambar 4.7 Data <i>Log IDS</i> Setelah Pengisian Variabel Global.....	32
Gambar 4.8 Data <i>Log IDS</i> Sebelum <i>Label Encoding</i>	34
Gambar 4.9 Data <i>Log IDS</i> Setelah <i>Label Encoding</i>	34
Gambar 4.10 Sample Data Latih Setelah <i>10-fold Cross Validation</i>	35
Gambar 4.11 Sample Data Uji Setelah <i>10-fold Cross Validation</i>	36
Gambar 4.12 Pohon Keputusan untuk Semua Fitur.....	37
Gambar 4.13 Contoh Aturan yang Telah Dibuat (Seluruh Fitur)	38
Gambar 4.14 Pohon Keputusan untuk Fitur Penting	45
Gambar 4.15 Contoh Aturan yang Telah Dibuat (Fitur Penting)	45
Gambar 4.16 <i>Timestamp</i> dengan <i>Destination Port</i>	46
Gambar 4.17 Source Country dengan Severity dan Category	47
Gambar 4.18 Peta Persebaran Penyerang (<i>Coordinate Map</i>)	47
Gambar 4.19 Peta Persebaran Penyerang (<i>Map</i>).....	48
Gambar 4.20 Severity dengan Destination Port dan Signature.....	48
Gambar 4.21 Severity dengan Destination Port dan Signature.....	49
Gambar 4.22 Destination Country	49
Gambar 4.23 Analisis Serangan.....	50

Gambar 4.24 Analisis Serangan..... 50

DAFTAR LAMPIRAN

Lampiran 1 Decision Tree Seluruh Fitur

Lampiran 2 Aturan Decision Tree Semua Fitur

Lampiran 3 Hasil Prediksi Semua Fitur

Lampiran 4 Decision Tree Fitur Penting

Lampiran 5 Aturan Decision Ttree Fitur Penting

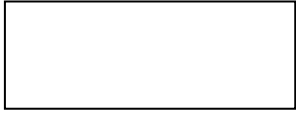
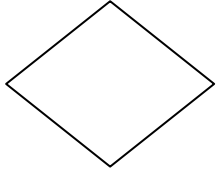



Lampiran 6 Kode Program Parsing Data

Lampiran 7 Kode Program Mengisi City dan Region

Lampiran 8 Kode Program Aturan Seleksi Fitur dan Klasifikasi

Lampiran 9 Similarity Index Skripsi

DAFTAR SIMBOL

Simbol	Nama Simbol	Keterangan
	Simbol Proses	Menggambarkan Proses
	Simbol Decision	Simbol pemilihan proses berdasarkan kondisi yang ada
	Simbol arah data atau arus data	Sebagai petunjuk arah data dan arus data pada proses
	Simbol Terminator	Simbol untuk permulaan atau akhir dari suatu kegiatan
	Simbol Input-Output	Simbol yang menyatakan proses input dan output tanpa tergantung jenis peralatannya