

**SISTEM ANALISIS DAN KLASIFIKASI RISIKO MALICIOUS TRAFFIC
PADA LOG HONEYPOT MENGGUNAKAN ALGORITMA RANDOM
FOREST**



FARHANI AMANAH
NIM. 2210511021

PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2026

**SISTEM ANALISIS DAN KLASIFIKASI RISIKO MALICIOUS TRAFFIC
PADA LOG HONEYPOT MENGGUNAKAN ALGORITMA RANDOM
FOREST**

**FARHANI AMANAH
NIM. 2210511021**

SKRIPSI

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

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
JAKARTA
2026**

LEMBAR PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

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

Nama : Farhani Amanah

NIM : 2210511021

Tanggal : 26 Juli 2026

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

Jakarta, 26 Juli 2026

Yang Menyatakan,



Farhani Amanah

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI

PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

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

Nama : Farhani Amanah

NIM : 2210511021

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 skripsi saya yang berjudul :

SISTEM ANALISIS DAN KLASIFIKASI RISIKO MALICIOUS TRAFFIC PADA LOG HONEYPOT MENGGUNAKAN ALGORITMA RANDOM FOREST

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/memformatkan, mengelola dalam bentuk pangkalan data (basis data), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Jakarta

Pada tanggal 26 Juli 2026

Yang Menyatakan,



Farhani Amanah

LEMBAR PERSETUJUAN SIDANG

LEMBAR PERSETUJUAN

Yang bertanda tangan di bawah ini:

Nama : Farhani Amanah
NIM : 2210511021
Program Studi : Informatika Program Sarjana/~~Sistem Informasi Program~~
~~Sarjana/Sains Data Program Sarjana/Sistem Informasi Program~~
~~Diploma~~ (*Coret yang tidak perlu)
Judul Tugas Akhir : Sistem Analisis dan Klasifikasi Risiko Malicious Traffic pada Log
Honeypot menggunakan Algoritma Random Forest

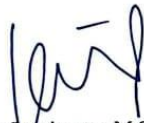
Dinyatakan telah memenuhi syarat dan menyetujui untuk mengikuti ujian sidang Tugas Akhir.

Jakarta, 2026

Menyetujui,

Dosen Pembimbing I,

Dosen Pembimbing II,



Prof. Dr. Ir. Supriyanto, M.Sc., IPM.
NIP. 197605082003121002



Kharisma Wiati Gusti, M.T.
NIP. 199006192022032007

Mengetahui,
Koordinator Program Studi,



Dr. Ridwan Raafi'udin, M.Kom.
NIP. 198805022021211001

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Judul : Sistem Analisis dan Klasifikasi Risiko Malicious Traffic pada Log
Honeypot Menggunakan Algoritma Random Forest
Nama : Farhani Amanah
NIM : 2210511021

Disetujui oleh :

Penguji 1:
Dr. Widya Cholil, M.I.T.



Penguji 2:
Anis Fitri Nur Masruriyah, S.Kom., M.Kom.



Pembimbing 1:
Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM.



Pembimbing 2:
Kharisma Wiati Gusti, S.T., M.T.



Diketahui oleh:

Koordinator Program Studi:
Dr. Ridwan Raafi'udin, S.Kom., M.Kom.
NIP. 198805022021211001

Dekan Fakultas Ilmu Komputer
Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM.
NIP. 197605082003121002



Tanggal Ujian Tugas Akhir
9 Juli 2026

ABSTRAK

Semakin meningkatnya ancaman siber mengakibatkan jumlah dan kompleksitas *malicious traffic* yang terekam di *honeypot* semakin tinggi sehingga diperlukan mekanisme analisis yang dapat mengidentifikasi tingkat risiko serangan secara terukur. Penelitian ini bertujuan membangun sistem analisis dan klasifikasi risiko *malicious traffic* pada log *honeypot* menggunakan algoritma *Random Forest*. Dataset yang digunakan berasal dari log *Honeypot* yang merekam aktivitas autentikasi dan interaksi penyerang. Tahapan penelitian meliputi *preprocessing data*, *feature engineering*, pembentukan label risiko menggunakan *pseudo labeling* berdasarkan *framework MITRE ATT&CK* dan *NIST SP 800-30*, penanganan ketidakseimbangan data menggunakan *Synthetic Minority Oversampling Technique (SMOTE)*, serta evaluasi model menggunakan *Stratified 5-Fold Cross Validation*. Hasil penelitian menunjukkan bahwa model *Random Forest* mampu mengklasifikasikan *malicious traffic* ke dalam kategori *Low Risk* dan *High Risk* dengan nilai *accuracy* sebesar 99,40%, *precision* sebesar 95,28%, *recall* sebesar 99,02%, *F1-score* sebesar 97,12%, dan *ROC-AUC* sebesar 99,95%. Selanjutnya model diimplementasikan ke dalam dashboard berbasis web untuk menampilkan hasil analisis, klasifikasi risiko, attack indication, dan security recommendation secara interaktif. Hasil penelitian menunjukkan bahwa kombinasi log *honeypot* dan *Random Forest* efektif digunakan untuk mendukung analisis risiko ancaman siber berdasarkan karakteristik perilaku serangan.

Kata kunci: *malicious traffic*, *honeypot*, *Random Forest*, klasifikasi risiko, keamanan siber

ABSTRACT

The increasing volume and sophistication of cyber threats have resulted in a growing amount of malicious traffic recorded by honeypots, creating the need for a mechanism capable of identifying attack risk levels in a measurable manner. This study aims to develop a risk analysis and classification system for malicious traffic based on honeypot logs using the Random Forest algorithm. The dataset was obtained from honeypot logs that record authentication activities and attacker interactions. The research process consisted of data preprocessing, feature engineering, risk label generation using pseudo labeling based on the MITRE ATT&CK framework and NIST SP 800-30, data balancing using the Synthetic Minority Oversampling Technique (SMOTE), and model evaluation through Stratified 5-Fold Cross Validation. The results show that the Random Forest model successfully classified malicious traffic into Low Risk and High Risk categories, achieving an accuracy of 99,40%, precision of 95,28%, recall of 99,02%, F1-score of 97,12%, and ROC-AUC of 99,95%. The model was subsequently implemented in a web-based dashboard to provide interactive analysis results, risk classification, attack indications, and security recommendations. These findings demonstrate that the combination of honeypot logs and Random Forest is effective in supporting cyber threat risk analysis based on attack behavior characteristics.

Keywords: malicious traffic, honeypot, Random Forest, risk classification, cybersecurity

DAFTAR ISI

LEMBAR PERNYATAAN ORISINALITAS	i
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI.....	ii
LEMBAR PERSETUJUAN SIDANG	iii
LEMBAR PENGESAHAN	iv
ABSTRAK	v
ABSTRACT.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xii
DAFTAR RUMUS.....	xiv
DAFTAR LAMPIRAN	xv
KATA PENGANTAR.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat Penelitian.....	4
1.4.1 Tujuan Penelitian	4
1.4.2 Manfaat Penelitian	4
1.5 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA	6
2.1 Kajian Teoretis.....	6
2.1.1 <i>Framework MITRE ATT&CK</i>	6
2.1.2 <i>Framework NIST SP 800 – 30</i>	7
2.1.3 Dasar Klasifikasi Risiko Berdasarkan <i>Framework MITRE</i> <i>ATT&CK</i> dan <i>NIST 800-30</i>	8
2.1.4 Sistem Website dan Dashboard.....	11

2.1.5 Keamanan Siber.....	12
2.1.6 <i>Malicious Traffic</i>	12
2.1.7 <i>Honeypot</i> sebagai Sumber Data Ancaman.....	13
2.1.8 <i>Random Forest</i>	15
2.1.9 <i>ADDIE Model</i>	17
2.1.10 <i>Confusion Matrix</i> dan Evaluasi Model	18
2.1.11 <i>Brute Force Attack</i>	20
2.1.12 <i>Port / Service Scanning</i>	21
2.1.13 <i>K-Fold Cross Validation</i>	22
2.1.14 <i>SMOTE</i>	23
2.1.15 <i>Feature Engineering</i>	24
2.1.16 <i>Pseudo Labeling</i>	25
2.2 Perumusan Hipotesis	26
2.3 Penelitian Terdahulu	26
BAB III METODE PENELITIAN.....	33
3.1 Jenis dan Pendekatan Penelitian.....	33
3.2 Objek dan Sumber Data Penelitian	35
3.3 Tahapan Penelitian Menggunakan <i>ADDIE</i>	35
3.3.1 <i>Analysis</i>	35
3.3.2 <i>Design</i>	37
3.3.3 <i>Development</i> (Pengembangan)	39
3.3.4 Evaluasi Model	52
3.3.5 <i>Implementation</i>	53
3.3.6 Evaluasi Sistem.....	54
3.4 Pengujian Hipotesis	55
3.5 Alat dan Bahan Penelitian	55
3.5.1 Alat Penelitian.....	55
3.5.2 Bahan Penelitian	56
BAB IV HASIL DAN PEMBAHASAN.....	58
4.1 Tahap <i>Analyze</i> (Analisis)	58
4.1.1 Hasil Identifikasi Masalah	58
4.1.2 Hasil Studi Literatur.....	59
4.1.3 Hasil Analisis Penelitian Sebelumnya	59

4.1.4 Hasil Analisis Dataset <i>Honeypot</i>	59
4.1.5 Hasil Analisis Kebutuhan Sistem.....	62
4.2 Tahap <i>Design</i> (Desain)	64
4.3 Tahap <i>Development</i> (Pengembangan)	65
4.3.1 Pengembangan Model <i>Random Forest</i>	65
4.3.2 Pengembangan Dashboard Sistem.....	89
4.4 Evaluasi Model.....	92
4.4.1 <i>Cross Validation</i>	92
4.4.2 Evaluasi Model <i>Random Forest</i>	95
4.4.3 Analisis <i>Feature Importance</i>	99
4.5 Tahap <i>Implementation</i> (implementasi)	102
4.6 Evaluasi Sistem	107
BAB V PENUTUP	111
5.1 Kesimpulan.....	111
5.2 Saran.....	111
DAFTAR PUSTAKA	113
LAMPIRAN.....	117

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur <i>Honeypot</i>	14
Gambar 2. 2 Arsitektur <i>Random Forest</i>	15
Gambar 2. 3 Tahapan Model <i>ADDIE</i>	17
Gambar 2. 4 Cara Kerja <i>K-Fold</i>	22
Gambar 3. 1 Alur Penelitian.....	34
Gambar 3. 2 Alur Tahap Analisis	35
Gambar 3. 3 Arsitektur Sistem Analisis dan Klasifikasi Risiko <i>Malicious Traffic</i>	38
Gambar 3. 4 Alur Pemrosesan Data	38
Gambar 3. 5 Arsitektur Model Klasifikasi Risiko <i>Malicious Traffic</i>	39
Gambar 3. 6 Alur Pengembangan Model <i>Random Forest</i>	40
Gambar 3. 7 Perbandingan <i>Random Forest</i> Standar dan Pengembangan Penelitian	41
Gambar 3. 8 Alur Pengembangan Dashboard Sistem	51
Gambar 3. 9 Alur Evaluasi Model.....	52
Gambar 3. 10 Alur Implementasi Sistem	53
Gambar 3. 11 Alur Evaluasi Sistem	54
Gambar 4. 1 Dataset.....	60
Gambar 4. 2 Data Sebelum <i>Preprocessing</i>	65
Gambar 4. 3 Hasil Pemeriksaan <i>Missing Value</i>	66
Gambar 4. 4 Sebelum dan Sesudah Dilakukan <i>Data Cleaning</i>	66
Gambar 4. 5 Sebelum dan Sesudah Standardisasi	67
Gambar 4. 6 Hasil Akhir <i>Preprocessing</i>	68
Gambar 4. 7 Dataset Sebelum <i>Feature Engineering</i>	68
Gambar 4. 8 Hasil Pemetaan Atribut Menjadi Fitur Numerik	70
Gambar 4. 9 Hasil Statistik Deskriptif.....	71
Gambar 4. 10 Contoh Proses Perhitungan <i>Risk Score</i> pada <i>Source IP</i>	72
Gambar 4. 11 Dataset Hasil <i>Pseudo Labeling</i>	73
Gambar 4. 12 Distribusi Hasil <i>Pseudo Labeling</i>	74
Gambar 4. 13 Daftar Fitur.....	75
Gambar 4. 14 Dataset Target.....	75
Gambar 4. 15 Cuplikan Data Latih	78

Gambar 4. 16 Cuplikan Data Uji.....	78
Gambar 4. 17 Grafik Perbandingan Jumlah Data Latih dan Data Uji	80
Gambar 4. 18 Grafik Distribusi Label Data Latih dan Data Uji	80
Gambar 4. 19 Cuplikan Data Latih dan Target Klasifikasi	82
Gambar 4. 20 Grafik Distribusi Kelas Sebelum <i>SMOTE</i>	83
Gambar 4. 21 Dataset Data Latih Setelah <i>SMOTE</i>	83
Gambar 4. 22 Grafik Distribusi Kelas Setelah <i>SMOTE</i>	84
Gambar 4. 23 Akurasi Setiap <i>Fold</i> dan Rata-Rata Seluruh Metrik Evaluasi	94
Gambar 4. 24 <i>Confusion Matrix</i>	96
Gambar 4. 25 <i>ROC Curve Random Forest</i>	98
Gambar 4. 26 <i>Feature Importance Random Forest</i>	100
Gambar 4. 27 Tahapan Implementasi Sistem.....	102
Gambar 4. 28 Dashboard Sistem Analisis dan Klasifikasi Risiko <i>Malicious Traffic</i>	103
Gambar 4. 29 Proses Analisis Dataset.....	104
Gambar 4. 30 Hasil Implementasi Dashboard	105
Gambar 4. 31 <i>Detection Result</i> Kategori <i>High Risk</i>	106
Gambar 4. 32 <i>Detection Result</i> Kategori <i>Low Risk</i>	106

DAFTAR TABEL

Tabel 2. 1 Dasar Klasifikasi Risiko berdasarkan <i>Framework MITRE ATT&CK</i> dan <i>NIST 800-30</i>	8
Tabel 2. 2 <i>Confusion Matrix</i>	18
Tabel 2. 3 Penelitian Terdahulu.....	26
Tabel 3. 1 Contoh Data sebelum Validasi	42
Tabel 3. 2 Contoh Data sesudah Validasi	42
Tabel 3. 3 Contoh Data sebelum <i>Data Cleaning</i>	43
Tabel 3. 4 Contoh Data sesudah <i>Data Cleaning</i>	43
Tabel 3. 5 Contoh Data sebelum Normalisasi.....	43
Tabel 3. 6 Contoh Data sesudah Normalisasi.....	44
Tabel 3. 7 Contoh Atribut sebelum Seleksi.....	44
Tabel 3. 8 Contoh Atribut sesudah Seleksi.....	45
Tabel 3. 9 Contoh Data sebelum Konversi Format	45
Tabel 3. 10 Contoh Data setelah Konversi Format	45
Tabel 3. 11 Contoh sebelum <i>Feature Engineering</i>	46
Tabel 3. 12 Contoh sesudah <i>Feature Engineering</i>	46
Tabel 3. 13 Contoh Fitur Rasio	47
Tabel 3. 14 Contoh <i>Pseudo Labeling</i>	47
Tabel 3. 15 Contoh Pembagian Data sebelum dan sesudah <i>SMOTE</i>	49
Tabel 3. 16 Contoh Distribusi Data pada tiap <i>Fold</i>	49
Tabel 3. 17 Perangkat Keras.....	55
Tabel 3. 18 Perangkat Lunak.....	56
Tabel 4. 1 Hasil Identifikasi Atribut Dataset.....	60
Tabel 4. 2 Hasil Seleksi Atribut Penelitian.....	61
Tabel 4. 3 Kebutuhan Fungsional Sistem.....	62
Tabel 4. 4 Kebutuhan Nonfungsional Sistem.....	63
Tabel 4. 5 Perbandingan Sebelum dan Sesudah <i>Data Cleaning</i>	67
Tabel 4. 6 Hubungan Atribut dan Fitur	69
Tabel 4. 7 Hasil Perhitungan <i>Threshold</i>	72
Tabel 4. 8 Daftar Fitur	75
Tabel 4. 9 Hasil Persiapan Data Pelatihan	76

Tabel 4. 10 Pembagian Data.....	77
Tabel 4. 11 Dimensi Dataset Setelah <i>Train-Test Split</i>	77
Tabel 4. 12 Distribusi Label Data Latih	79
Tabel 4. 13 Distribusi Label Data Uji	79
Tabel 4. 14 Distribusi Kelas Sebelum <i>SMOTE</i>	82
Tabel 4. 15 Distribusi Kelas Setelah <i>SMOTE</i>	84
Tabel 4. 16 Perbandingan Distribusi Kelas Sebelum dan Sesudah <i>SMOTE</i>	85
Tabel 4. 17 Ringkasan Hasil <i>SMOTE</i>	85
Tabel 4. 18 Perbandingan Metrik Sebelum dan Sesudah <i>SMOTE</i>	86
Tabel 4. 19 Parameter Algoritma <i>Random Forest</i>	87
Tabel 4. 20 Alur Pipeline Pelatihan Model	87
Tabel 4. 21 Implementasi Modul Pembacaan Dataset	89
Tabel 4. 22 Implementasi Modul Pemrosesan Data.....	90
Tabel 4. 23 Implementasi Modul Klasifikasi Risiko.....	91
Tabel 4. 24 Implementasi Model Penyajian Informasi	92
Tabel 4. 25 Parameter <i>Cross Validation</i>	93
Tabel 4. 26 Hasil <i>Cross Validation</i> Setiap <i>Fold</i>	93
Tabel 4. 27 Rata-rata Hasil <i>Cross Validation</i>	93
Tabel 4. 28 Hasil <i>Classification Report</i>	95
Tabel 4. 29 Hasil <i>Confusion Matrix</i>	96
Tabel 4. 30 Hasil Evaluasi Model	97
Tabel 4. 31 Nilai <i>Feature Importance</i>	100
Tabel 4. 32 Hasil <i>Black Box Testing</i>	107
Tabel 4. 33 Hasil Pengujian Integrasi Sistem.....	108
Tabel 4. 34 Hasil <i>Performance Testing</i>	109

DAFTAR RUMUS

Rumus 2. 1	16
Rumus 2. 2	16
Rumus 2. 3	16
Rumus 2. 4	16
Rumus 2. 5	19
Rumus 2. 6	19
Rumus 2. 7	19
Rumus 2. 8	20

DAFTAR LAMPIRAN

Lampiran 1. Lembar Validasi Data oleh Pakar	117
Lampiran 2. Kode Program requirements.txt.....	122
Lampiran 3. Kode Program Load Dataset.....	122
Lampiran 4. Kode Program Preprocessing Data.....	122
Lampiran 5. Kode Program Feature Engineering	123
Lampiran 6. Kode Program Pseudo Labeling	124
Lampiran 7. Kode Program Pembentukan Data Training.....	124
Lampiran 8. Kode Program Train Test Split	125
Lampiran 9. Kode Program Penyeimbangan Data Menggunakan SMOTE	125
Lampiran 10. Kode Program Pelatihan Model Random Forest.....	125
Lampiran 11. Kode Program Cross Validation	126
Lampiran 12. Kode Program Evaluasi Model.....	126
Lampiran 13. Kode Program Analisis Feature Importance.....	127
Lampiran 14. Kode Program Penyimpanan Model.....	127
Lampiran 15. Kode Program Backend app.py	128
Lampiran 16. Kode Program Frontend dashboard.html.....	130
Lampiran 17. Hasil Turnitin.....	142
Lampiran 18 Daftar Riwayat Hidup.....	143

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat-Nya sehingga Tugas Akhir ini berhasil diselesaikan. Tugas Akhir ini dilakukan dalam bentuk Skripsi. Tugas Akhir ini dilakukan sejak bulan September 2025 sampai bulan April 2026 dengan judul “Sistem Analisis dan Klasifikasi Risiko *Malicious Traffic* Pada Log *Honeypot* menggunakan Algoritma *Random Forest*”.

Terima kasih penulis ucapkan kepada:

1. Bapak Hasan Basri dan Ibu Fatimah Zahara selaku Ayah dan Bunda dari penulis yang senantiasa memberikan *effort* yang sangat luar biasa kepada penulis, dukungan moral, doa, serta motivasi yang tiada henti.
2. Bapak Prof. Dr. Ir. Supriyanto, M.Sc., IPM. selaku Dekan Fakultas Ilmu Komputer sekaligus dosen pembimbing, dan Ibu Kharisma Wiati Gusti, M.T. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan saran kepada penulis selama proses penyusunan skripsi ini.
3. Bapak Dr. Ridwan Raafi'udin, S.Kom, M.Kom. selaku Ketua Program Studi Informatika.
4. Bapak Basuki Erwin Setiyadi, S.S.T.P., M.Si. selaku narasumber dari Direktorat Operasi Keamanan Siber (D21) Badan Siber dan Sandi Negara (BSSN), serta Badan Siber dan Sandi Negara (BSSN) yang telah memberikan tanggapan, informasi, dan jawaban atas pertanyaan penelitian yang diajukan penulis sehingga dapat mendukung penyusunan skripsi ini.

Semoga Tugas Akhir ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Jakarta, 26 Juli 2026

Farhani Amanah