

**PENGEMBANGAN MODEL MACHINE LEARNING UNTUK
DETEKSI ANOMALI DAN SERANGAN PADA JARINGAN
KOMPUTER**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 INFORMATIKA



disusun oleh

ERMAN SAPUTRA

21.11.3859

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PENGEMBANGAN MODEL MACHINE LEARNING UNTUK
DETEKSI ANOMALY DAN SERANGAN PADA JARINGAN
KOMPUTER**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 INFORMATIKA



disusun oleh

ERMAN SAPUTRA

21.11.3859

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PENGEMBANGAN MODEL MACHINE LEARNING UNTUK
DETEKSI ANOMALY DAN SERANGAN PADA JARINGAN
KOMPUTER**

yang disusun dan diajukan oleh

Nama Mahasiswa

Erman Saputra

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Juli 2025

Dosen Pembimbing,

Hanafi, S.Kom., M.Eng., Ph.D.
NIK. 190302024

HALAMAN PENGESAHAN

SKRIPSI

PENGEMBANGAN MODEL MACHINE LEARNING UNTUK
DETEKSI ANOMALY DAN SERANGAN PADA JARINGAN
KOMPUTER

yang disusun dan diajukan oleh

Nama Mahasiswa

Erman Saputra

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Juli 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Bayu Setiaji, M.Kom.
NIK. 190302216

Bambang Pulu Hartato S.Kom., M.Eng.
NIK. 190302707

Hanafi, S.Kom., M.Eng., Ph.D.D
NIK. 190302024



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Juli 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Erman Saputra
NIM : 21.11.3859

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Hanafi, S.Kom., M.Eng., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Juli 2025

Yang Menyatakan,



Nama Mahasiswa

HALAMAN PERSEMBAHAN KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan ke hadirat Allah SWT, karena atas limpahan rahmat, taufik, dan hidayah-Nya, penulis dapat menyelesaikan skripsi yang berjudul: “PENGEMBANGAN MODEL MACHINE LEARNING UNTUK DETEKSI ANOMALI DAN SERANGAN PADA JARINGAN KOMPUTER”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik Komputer pada Program Studi Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Ucapan terima kasih penulis:

1. Allah SWT, atas segala nikmat, kemudahan, dan kekuatan yang diberikan selama proses penyusunan skripsi ini.
2. Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
3. Hanafi, S.Kom., M.Eng., Ph.D, selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan yang sangat berarti selama proses pengerjaan skripsi.
4. Kedua orang tua dan keluarga tercinta, atas segala doa, dukungan moral dan materi, serta semangat yang selalu menyertai setiap langkah penulis
5. Teman-teman terdekat: Cahyo Pamungkas, Purnama Sari, Romiza Farrel, dan Kevin , atas motivasi, bantuan, dan kebersamaannya selama proses ini.
6. Teman-teman IF01, atas kebersamaan, motivasi, dan bantuan yang diberikan selama masa studi dan penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan. Semoga karya ini dapat memberikan manfaat, khususnya dalam bidang keamanan jaringan komputer.

Yogyakarta, 16 Juli 2025

Penulis

DAFTAR ISI

(gunakan tools table of content pada menu references di Word)

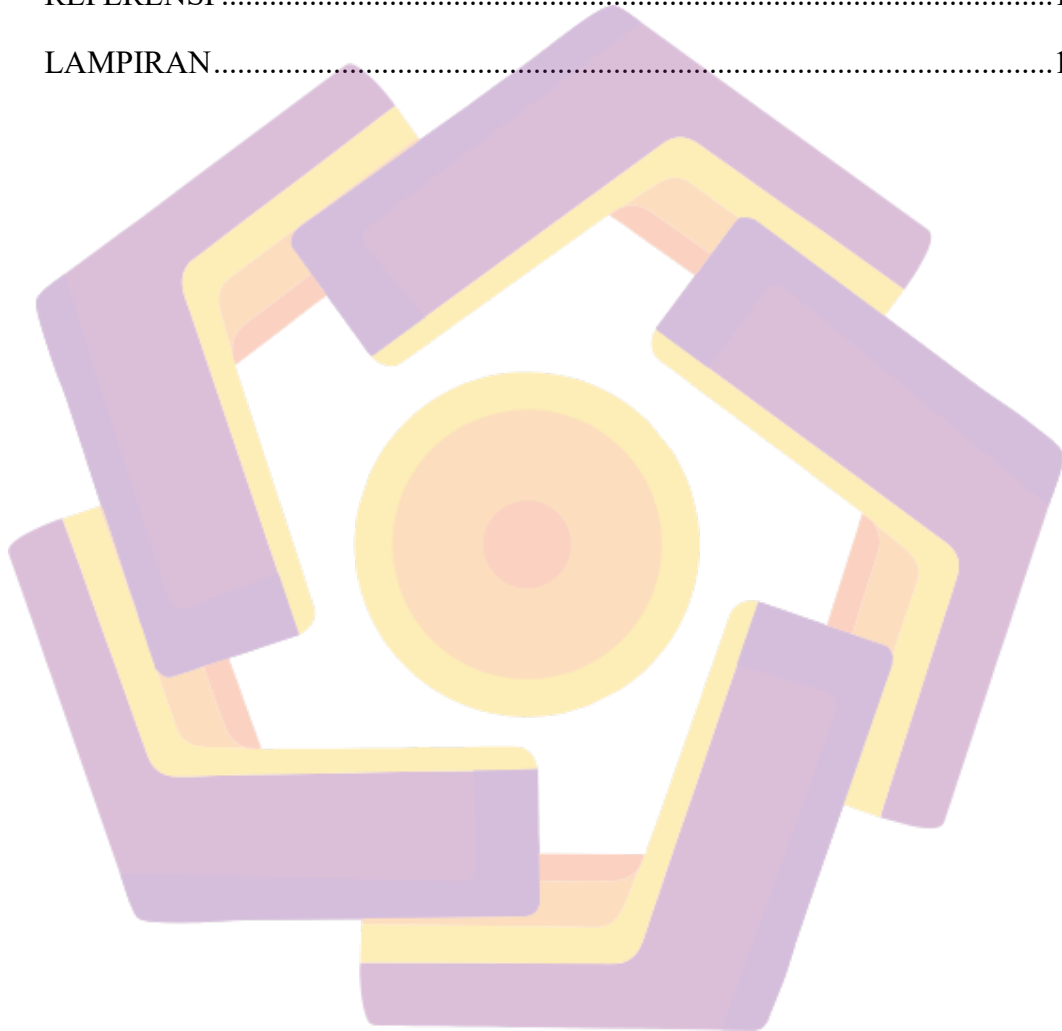
Daftar Isi

HALAMAN JUDUL	1
HALAMAN PERSETUJUAN.....	2
HALAMAN PENGESAHAN	3
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	4
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	5
DAFTAR ISI.....	6
DAFTAR TABEL.....	10
DAFTAR GAMBAR.....	11
DAFTAR LAMPIRAN.....	13
DAFTAR LAMBANG DAN SINGKATAN	14
DAFTAR ISTILAH.....	15
INTISARI	16
<i>ABSTRACT</i>	17
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4

BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	12
2.2.1 Jaringan Komputer	12
2.2.2 Keamanan Siber	12
2.2.3 Serangan Siber	12
2.2.4 <i>Machine Learning</i>	17
BAB III METODE PENELITIAN	20
3.1 Objek Penelitian	20
3.2 Alur Penelitian	21
3.2.1 Gambaran Umum Data	22
3.2.2 Pra-pemrosesan data	24
3.2.3 Pembagian Data	27
3.2.4 SMOTETomek.....	27
3.2.5 Pemilihan Algoritma Klasifikasi.....	28
3.2.6 Pemodelan Awal	32
3.2.7 GridSearch CV.....	33
3.2.8 Evaluasi.....	34
3.3 Alat dan Bahan.....	36
3.3.1 Data Penelitian	36
3.3.2 Alat.....	37
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Gambaran Dataset.....	40
4.1.1 NSL-KDD	40
4.1.2 UNSW-NB15	41

4.2	Pra-Pemrosesan Data	42
4.2.1	Eksplorasi Data	42
4.2.2	Pengubahan Label	43
4.2.3	Visualisasi Kategori	44
4.2.4	Visualisasi Ketidakseimbangan Data	48
4.2.5	Konversi Tipe Data Kategorik	50
4.2.6	Analisis Korelasi Fitur	50
4.2.7	Pembersihan Data	52
4.2.8	Pembentukan Set Fitur dan Target	53
4.3	Pembagian Data	53
4.4	SMOTETomek	55
4.5	Pemodelan Awal	57
4.5.1	<i>Logistic Regression</i>	58
4.5.2	Random Forest	58
4.5.3	Support Vector Machine (SVM)	59
4.5.4	Naïve Bayes	59
4.5.6	K-Nearest Neighbors (KNN)	60
4.6	<i>GridSearch CV</i>	60
4.6.1	Gridsearch CV <i>Logistic Regression</i>	61
4.6.2	Gridsearch CV Random Forest	62
4.6.4	Gridsearch CV Naïve Bayes	64
4.6.5	Gridsearch CV K-Nearest Neighbors (KNN)	66
4.7	Evaluasi	67
4.7.1	Perbandingan Pemodelan Awal	67
4.7.2	Perbandingan Setelah GridsearchCV	95

4.7.3 Analisis Komparatif dan Penentuan Model Terbaik.....	130
BAB V PENUTUP	132
5.1 Kesimpulan	132
5.2 Saran	133
REFERENSI	134
LAMPIRAN.....	140



DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 3.1 Dataset NSL-KDD	22
Tabel 3.2 Dataset UNSW-NB15	23
Tabel 3.3 Parameter GridSearch CV	33
Tabel 4.1 Fitur Kategorik Dataset	42
Tabel 4.2 Label Klasifikasi	43
Tabel 4.3 Fitur Yang Di Hapus	52
Tabel 4. 4 Laporan Klasifikasi Biner Logistic Regression	68
Tabel 4.5 Laporan Klasifikasi Biner Random Forest	70
Tabel 4.6 Laporan Klasifikasi Biner Support Vector Machine	72
Tabel 4. 7 Laporan Klasifikasi Biner Naive Bayes	74
Tabel 4. 8 Laporan Klasifikasi Biner K-Nearest Neighbors	76
Tabel 4.9 Laporan Klasifikasi Multi Logistic Regression	79
Tabel 4.10 Laporan Klasifikasi Multi Random Forest	82
Tabel 4.11 Laporan Klasifikasi Multi Support Vector Machine	84
Tabel 4.12 Laporan Klasifikasi Multi Naive Bayes	87
Tabel 4. 13 Laporan Klasifikasi Multi K-Nearest Neighbors	89
Tabel 4.14 Laporan Klasifikasi Biner Logistic Regression Gridsearch CV	96
Tabel 4.15 Laporan Klasifikasi Biner Random Forest Gridsearch CV	99
Tabel 4.16 Laporan Klasifikasi Biner SVM Gridsearch CV	102
Tabel 4.17 Laporan Klasifikasi Biner Naive Bayes Gridsearch CV	105
Tabel 4. 18 Laporan Klasifikasi Biner KNN Gridsearch CV	107
Tabel 4.19 Laporan Klasifikasi Multi Logistic Regression Gridsearch CV	111
Tabel 4.20 Laporan Klasifikasi Multi Random Forest Gridsearch CV	114
Tabel 4.21 Laporan Klasifikasi Multi SVM Gridsearch CV	118
Tabel 4.22 Laporan Klasifikasi Multi Naive Bayes Gridsearch CV	120
Tabel 4.23 Laporan Klasifikasi Multi KNN Gridsearch CV	124

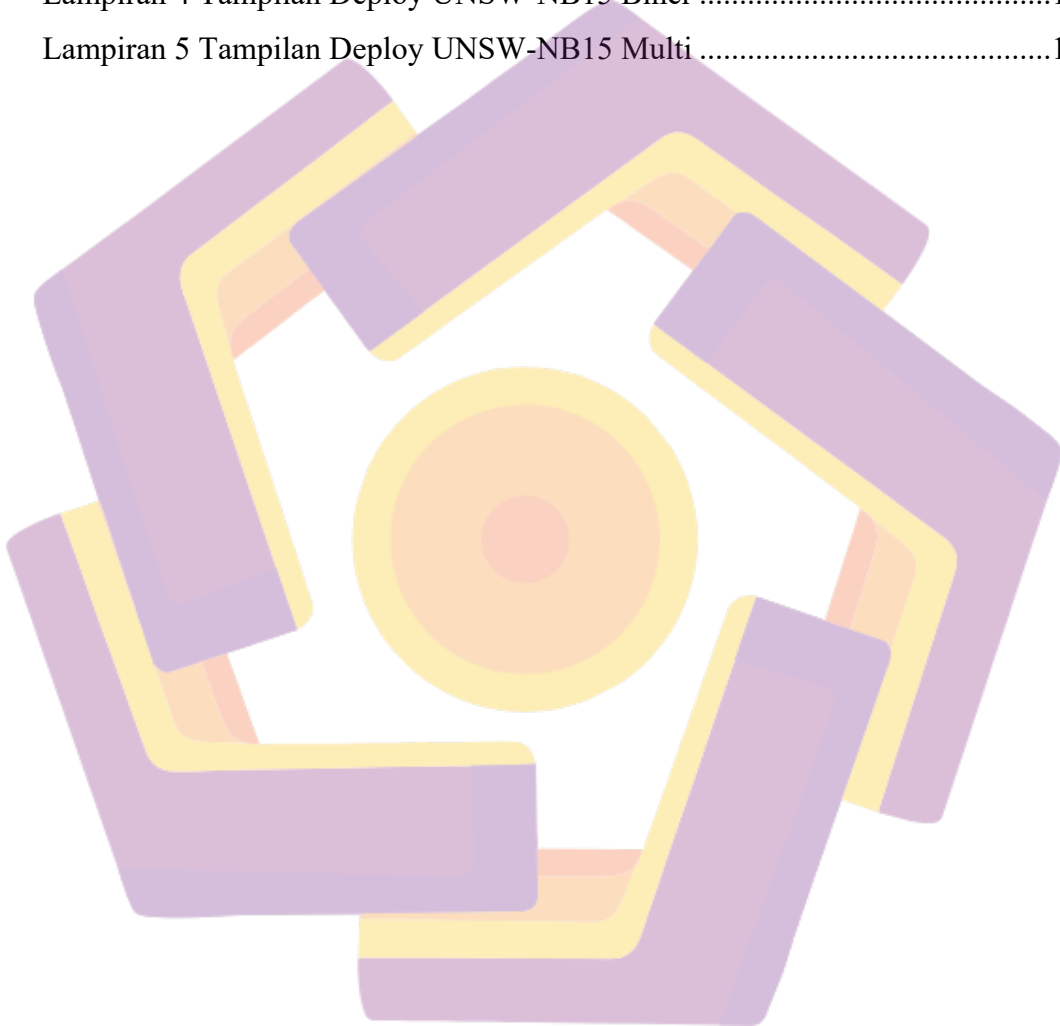
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	21
Gambar 3.2 Visualisasi SMOTETomek	28
Gambar 3. 3 Confusion Matrix	35
Gambar 4.1 protocol_type biner NSL-KDD	44
Gambar 4.2 flag biner NSL-KDD	44
Gambar 4.3 Service Biner NSL-KDD	45
Gambar 4.4 protocol_type Multi NSL-KDD	45
Gambar 4. 5 Flag Multi NSL-KDD	45
Gambar 4.6 Service Multi NSL-KDD	46
Gambar 4.7 Proto Biner UNSW-NB15	46
Gambar 4.8 Service Biner UNSW-NB15	46
Gambar 4.9 State Biner UNSW-NB15	47
Gambar 4.10 Proto Multi UNSW-NB15	47
Gambar 4.11 Service Multi UNSW-NB15	47
Gambar 4.12 State Multi UNSW-NB15	47
Gambar 4.13 Label Biner NSL-KDD	48
Gambar 4.14 Label Multi NSL-KDD	49
Gambar 4.15 Label Biner UNSW-NB15	49
Gambar 4. 16 Label Multi UNSW-NB15	49
Gambar 4.17 Korelasi Fitur NSL-KDD	51
Gambar 4.18 Korelasi Fitur UNSW-NB15	51
Gambar 4.19 Data Train VS Testing Biner NSL-KDD	54
Gambar 4.20 Data Train VS Testing Multi NSL-KDD	54
Gambar 4.21 Data Train VS Testing Biner UNSW-NB15	54
Gambar 4.22 Data Train VS Testing Biner UNSW-NB15	55
Gambar 4.23 Label Biner NSL-KDD SMOTETomek	56
Gambar 4.24 Label Multi NSL-KDD SMOTETomek	56
Gambar 4.25 Label Biner UNSW-NB15 SMOTETomek	57
Gambar 4.26 Label Multi UNSW-NB15 SMOTETomek	57

Gambar 4.27 Confusion Matrix Klasifikasi Biner Logistic Regression	69
Gambar 4. 28 Confusion Matrix Klasifikasi Biner Random Forest	71
Gambar 4. 29 Confusion Matrix Klasifikasi Biner Support Vector Machine	73
Gambar 4.30 Confusion Matrix Klasifikasi Biner Naive Bayes.....	75
Gambar 4.31 Confusion Matrix Klasifikasi Biner K-Nearest-Neighbors	77
Gambar 4. 32 Confusion Matrix Klasifikasi Multi Logistic Regression	80
Gambar 4.33 Confusion Matrix Klasifikasi Multi Random Forest	83
Gambar 4.34 Confusion Matrix Klasifikasi Multi Support Vector Machine	85
Gambar 4.35 Confusion Matrix Klasifikasi Multi Naive Bayes.....	88
Gambar 4.36 Confusion Matrix Klasifikasi Multi K-Nearest Neighbors.....	91
Gambar 4.37 Laporan Klasifikasi Biner Pada Seluruh Model.....	93
Gambar 4.38 Laporan Klasifikasi Multi Pada Seluruh Model.....	94
Gambar 4.39 Confusion Matrix Klasifikasi Biner LR GridSearchCV	98
Gambar 4.40 Confusion Matrix Klasifikasi Biner RF GridSearchCV	101
Gambar 4.41 Confusion Matrix Klasifikasi Biner SVM GridSearchCV	103
Gambar 4.42 Confusion Matrix Klasifikasi Biner NB GridSearchCV.....	106
Gambar 4.43 Confusion Matrix Klasifikasi Biner KNN GridSearchCV	109
Gambar 4.44 Confusion Matrix Klasifikasi Multi LR GridSearchCV	112
Gambar 4. 45 Confusion Matrix Klasifikasi Multi RF GridSearchCV	116
Gambar 4. 46 Confusion Matrix Klasifikasi Multi SVM GridSearchCV	119
Gambar 4. 47 Confusion Matrix Klasifikasi Multi NB GridSearchCV.....	122
Gambar 4. 48 Confusion Matrix Klasifikasi Multi KNN GridSearchCV	125
Gambar 4.49 Laporan Klasifikasi Biner GridSearchCV Pada Seluruh Model....	127
Gambar 4.50 Laporan Klasifikasi Multi GridSearchCV Pada Seluruh Model....	128

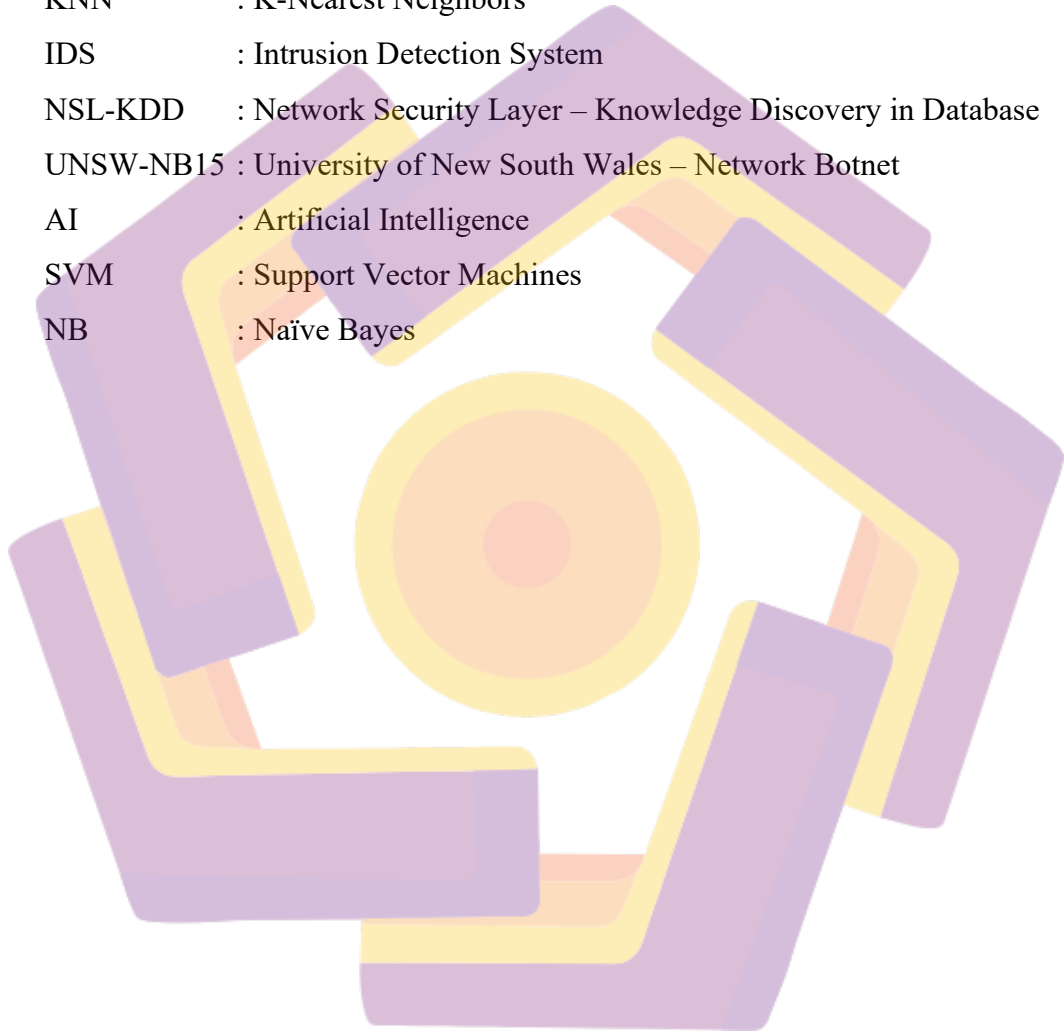
DAFTAR LAMPIRAN

Lampiran 1 Link Github	140
Lampiran 2 Tampilan Deploy NSL-KDD Biner	140
Lampiran 3 Tampilan Deploy NSL-KDD Multi	141
Lampiran 4 Tampilan Deploy UNSW-NB15 Biner	141
Lampiran 5 Tampilan Deploy UNSW-NB15 Multi	142

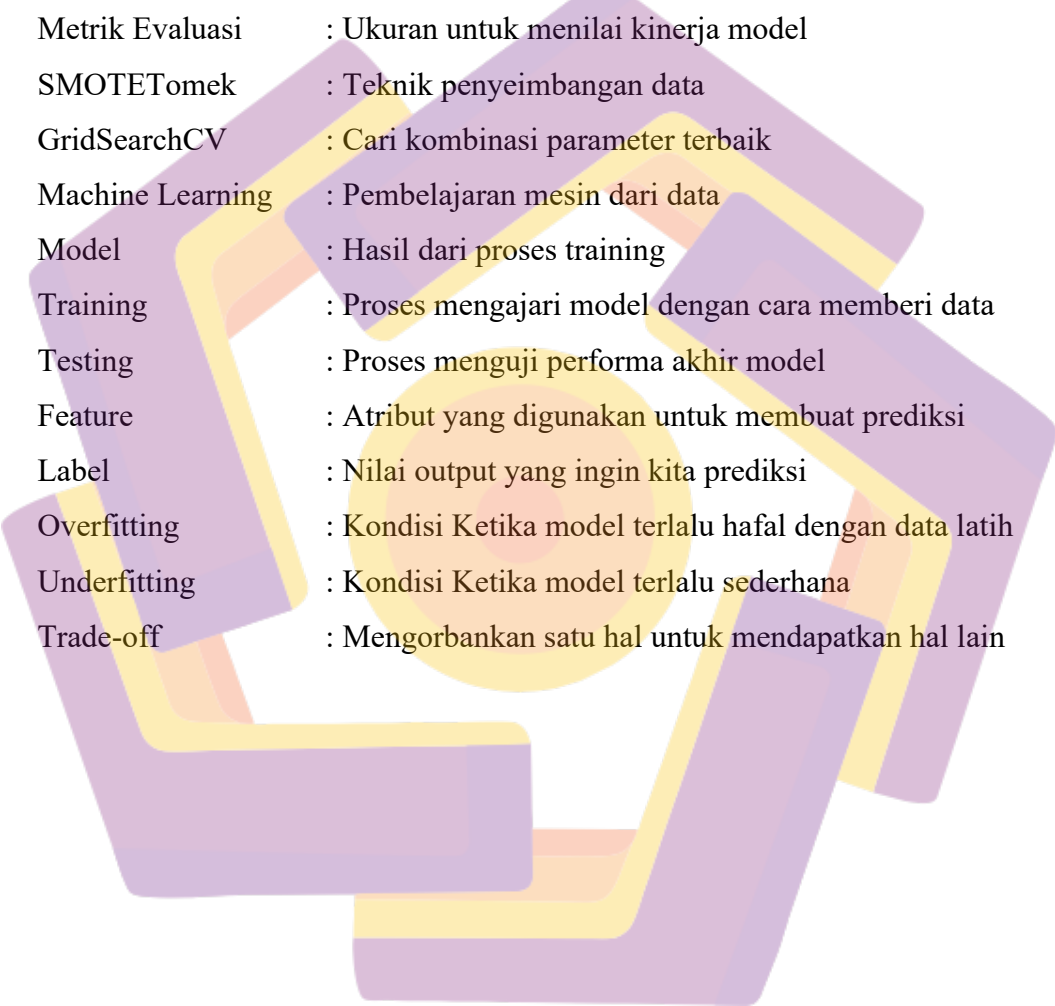


DAFTAR LAMBANG DAN SINGKATAN

ML	: Machine Learning
LR	: Logistic Regression
RF	: Random Forest
KNN	: K-Nearest Neighbors
IDS	: Intrusion Detection System
NSL-KDD	: Network Security Layer – Knowledge Discovery in Database
UNSW-NB15	: University of New South Wales – Network Botnet
AI	: Artificial Intelligence
SVM	: Support Vector Machines
NB	: Naïve Bayes



DAFTAR ISTILAH



Pearson Correlation	: Mengukur hubungan dua variable
Dataset	: Kumpulan semua data yang digunakan
Preprocessing	: Tahap persiapan data seblum masuk ke model
Cross-Validation	: Teknik uji model yang lebih andal untuk performa umum
Metrik Evaluasi	: Ukuran untuk menilai kinerja model
SMOTETomek	: Teknik penyeimbangan data
GridSearchCV	: Cari kombinasi parameter terbaik
Machine Learning	: Pembelajaran mesin dari data
Model	: Hasil dari proses training
Training	: Proses mengajari model dengan cara memberi data
Testing	: Proses menguji performa akhir model
Feature	: Atribut yang digunakan untuk membuat prediksi
Label	: Nilai output yang ingin kita prediksi
Overfitting	: Kondisi Ketika model terlalu hafal dengan data latih
Underfitting	: Kondisi Ketika model terlalu sederhana
Trade-off	: Mengorbankan satu hal untuk mendapatkan hal lain

INTISARI

Keamanan jaringan komputer semakin penting seiring dengan meningkatnya jumlah perangkat dan kompleksitas infrastruktur jaringan. Serangan seperti *Denial of Service (DoS)*, *Remote to Local (R2L)*, *User to Root (U2R)*, dan *Probe* menjadi ancaman serius yang perlu dideteksi sedini mungkin. Metode deteksi tradisional berbasis *signature* memiliki keterbatasan dalam menghadapi serangan baru yang belum terdokumentasi. Oleh karena itu, pendekatan berbasis *heuristic-based* digunakan untuk mendeteksi anomali secara proaktif.

Penelitian ini menggunakan lima algoritma *Machine Learning*, yaitu *Logistic Regression*, *Random Forest*, *Support Vector Machine (SVM)*, *Naïve Bayes*, dan *K-Nearest Neighbors (KNN)* untuk mengidentifikasi serangan siber. Dua dataset yang digunakan adalah *NSL-KDD* dan *UNSW-NB15* yang mencerminkan berbagai jenis serangan dan lalu lintas jaringan. Proses penelitian mencakup pra-pemrosesan data, pembagian data, penyeimbangan data menggunakan *SMOTETomek*, pelatihan model, dan optimasi dengan *GridSearchCV*.

Pada *NSL-KDD* biner, Akurasi *Random Forest* mencapai 100%, *SVM* 99%, *KNN* 100% dengan recall 99%, *Logistic Regression* 95%, dan *Naive Bayes* 89%. Pada *UNSW-NB15* biner, *Random Forest* 99%, *KNN* 97%, *SVM* 96%, *Logistic Regression* 95%, dan *Naive Bayes* 90%. Untuk multi-kelas *NSL-KDD*, Akurasi *Random Forest* 100%, *SVM* 99%, *KNN* 99%, *Logistic Regression* 94%, dan *Naive Bayes* 74%. Pada multi-kelas *UNSW-NB15*, *Random Forest* 85%, *KNN* 81%, *SVM* 77%, *Logistic Regression* 73%, dan *Naive Bayes* 60%. Hasil ini menunjukkan *Random Forest* sebagai model paling andal untuk pengembangan sistem deteksi dini serangan jaringan.

Kata kunci: Machine Learning, Deteksi Anomali, Serangan Siber, *NSL-KDD*, *UNSW-NB15*

ABSTRACT

The importance of computer network security is increasing along with the rise of connected devices and network complexity. Attacks such as Denial of Service (DoS), Remote to Local (R2L), User to Root (U2R), and Probe pose serious threats that must be detected early. Traditional signature-based detection methods are limited in identifying new and undocumented threats. Therefore, a heuristic-based approach is proposed to detect anomalies proactively.

This study employs five Machine Learning algorithms: Support Vector Machine (SVM), Random Forest, Naïve Bayes, K-Nearest Neighbors (KNN), and Logistic Regression to identify cyber attacks. Two datasets, NSL-KDD and UNSW-NB15, are used due to their variety of attack types and network traffic representation. The methodology includes data preprocessing, data splitting, class balancing with SMOTETomek, model training, and optimization using GridSearchCV.

In NSL-KDD binary, accuracy Random Forest achieved 100%, SVM 99%, KNN 100% with 99% recall, Logistic Regression 95%, and Naive Bayes 89%. In UNSW-NB15 binary, Random Forest reached 99%, KNN 97%, SVM 96%, Logistic Regression 95%, and Naive Bayes 90%. For NSL-KDD multi-class, Random Forest obtained 100%, SVM 99%, KNN 99%, Logistic Regression 94%, and Naive Bayes 74%. In UNSW-NB15 multi-class, accuracy Random Forest scored 85%, KNN 81%, SVM 77%, Logistic Regression 73%, and Naive Bayes 60%. These results highlight Random Forest as the most reliable model for developing early intrusion detection systems.

Keyword: Machine Learning, Anomaly Detection, Cyber Attack, NSL-KDD, UNSW-NB15