

**MODEL *WEB APPLICATION FIREWALL* BERBASIS *MACHINE*
LEARNING UNTUK MENCEGAH SERANGAN SIBER**

TUGAS AKHIR

NO. 1316/WM.FT.H6/T.ILKOM/TA/2025

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



Disusun oleh:

JOSERAY ARIMATEIA LOPES DA CRUZ

23122196

PROGRAM STUDI ILMU KOMPUTER

FAKULTAS TEKNIK

UNIVERSITAS KATOLIK WIDYA MANDIRA KUPANG

2026

HALAMAN PERSETUJUAN

TUGAS AKHIR

NO. 1316/WM.FT.H6/T.ILKOM/TA/2025

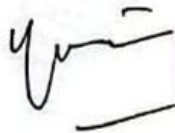
**MODEL *WEB APPLICATION FIREWALL* BERBASIS *MACHINE*
LEARNING UNTUK MENCEGAH SERANGAN SIBER**

JOSERAY ARIMATEIA LOPES DA CRUZ

23122196

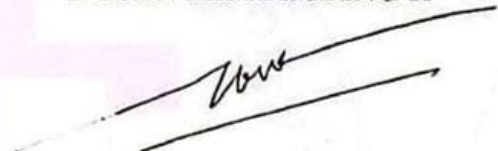
TELAH DIPERIKSA/DISETUJUI OLEH PEMBIMBING:

DOSEN PEMBIMBING I



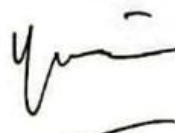
Yulianti Paula Bria, S.T., M.T., Ph.D
NIDN: 0823078702

DOSEN PEMBIMBING II



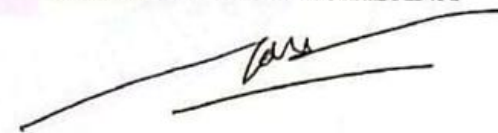
Donatus J. Manehat, S.Si., M.Kom
NIDN: 0828126601

KETUA PELAKSANA



Yulianti Paula Bria, S.T., M.T., Ph.D
NIDN: 0823078702

SEKRETARIS PELAKSANA



Donatus J. Manehat, S.Si., M.Kom
NIDN: 0828126601

HALAMAN PENGESAHAN

TUGAS AKHIR
NO. 1316/WM.FT.H6/T.ILKOM/TA/2025
MODEL *WEB APPLICATION FIREWALL* BERBASIS *MACHINE*
***LEARNING* UNTUK MENCEGAH SERANGAN SIBER**

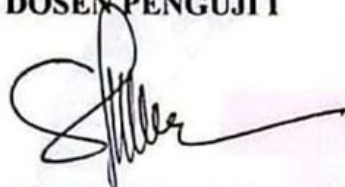
OLEH:

JOSERAY ARIMATEIA LOPES DA CRUZ
23122196

TELAH DIPERTAHANKAN DI DEPAN PENGUJI:

DI : KUPANG
PADA TANGGAL : 20 JANUARI 2026

DOSEN PENGUJI I



Sisilia D. Bakka Mau, S.Kom., M.T
NIDN: 0807098502

DOSEN PENGUJI II



Paul F. M. Tengangatu, S.Kom. M.T.I
NUPTK: 4955769670130332

DOSEN PENGUJI III



Yulianti Paula Bria, S.T., M.T., Ph.D
NIDN: 0823078702

MENGETAHUI

KETUA PROGRAM STUDI ILMU
KOMPUTER UNIKA WIDYA
MANDIRA KUPANG



Yovinia C. Hoar Siki, S.T., M.T
NIDN: 0805058803

MENGESAHKAN

DEKAN FAKULTAS TEKNIK
UNIKA WIDYA MANDIRA KUPANG



Dr. Don C. N. Da Costa, S.T., M.T
NIDN: 0820036801

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur, haru, tawa, tangis, dan segala warna emosi yang telah menemani perjalanan ini, Tugas Akhir ini dipersembahkan kepada:

1. Tuhan Yang Maha Esa, yang telah menuntun, membimbing, dan memberkati saya sehingga saya bisa yakin dan kuat untuk terus berusaha dan sampai di tahap ini.
2. Papa dan Mama, yang telah membesarkan saya, memberikan begitu banyak bantuan, cinta, perhatian, dukungan, dan segala hal lainnya yang tidak bisa saya sebutkan satu per satu.
3. Dosen Pembimbing I, Ibu Yulianti Paula Bria, S.T., M.T., Ph.D., dan Dosen Pembimbing II, Bapak Donatus Joseph Manchat, S.Si., M.Kom., atas waktu, perhatian, dan bantuan yang telah diberikan kepada saya hingga saat ini.
4. Para Dosen dan Tenaga Pendidik, atas banyaknya ilmu dan pengajaran, nasihat, dan bimbingan yang saya terima selama berkuliah dari awal semester hingga tahap ini.
5. Kakak-kakak saya yang selalu mendukung dan mendoakan saya.
6. Pacar saya, Sесilia Hingi Novita Sari Hokon, yang selalu memberikan dukungan, bantuan, dan waktunya dari awal pertemuan hingga saat ini.
7. *Os Seis*, yang telah menemani hari-hari perjalanan perkuliahan saya.
8. Teman-teman Kelas A Angkatan 2022 Program Studi Ilmu Komputer yang telah mendukung dan membantu saya hingga saat ini.
9. Semua pihak yang tidak dapat saya sebutkan satu per satu.

MOTTO

Tidak ada hal yang mustahil dicapai.

Kita hanya perlu usaha yang tepat di waktu yang tepat.

PERNYATAAN DAN KEASLIAN HASIL KARYA

Yang bertanda tangan di bawah ini:

Nama : Joseray Arimateia Lopes Da Cruz

NIM : 23122196

Fakultas/Prodi : Teknik/Illmu Komputer

Dengan ini, saya menyatakan bahwa Tugas Akhir ini yang berjudul “**Model Web Application Firewall Berbasis Machine Learning untuk Mencegah Serangan Siber**” adalah hasil karya saya sendiri. Karya ini disusun tanpa melakukan plagiarisme dan sepenuhnya berdasarkan penelitian, analisis, serta pemikiran yang telah saya lakukan. Apabila di kemudian hari, terbukti bahwa karya ini merupakan hasil plagiarisme atau terdapat pelanggaran akademik lainnya, saya bersedia menerima sanksi sesuai dengan peraturan yang berlaku di Universitas Katolik Widya Mandira Kupang. Demikian pernyataan ini saya buat dengan sebenar-benarnya dan penuh tanggung jawab.

Kupang, Januari 2026

Disahkan/Diketahui

Pembimbing



Yulianti Paula Bria, S.T., M.T., Ph.D
NIDN: 0823078702

Mahasiswa/Pemilik




Joseray Arimateia Lopes Da Cruz

KATA PENGANTAR

Puji syukur dihaturkan kepada Tuhan Yang Maha Esa karena atas berkat, rahmat, dan karunia-Nya yang besar, Tugas Akhir penelitian yang berjudul *Model Web Application Firewall Berbasis Machine Learning* untuk Mencegah Serangan Siber ini dapat diselesaikan. Tugas Akhir ini disusun untuk diajukan sebagai salah satu syarat kelulusan pada program Strata-1 di Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Katolik Widya Mandira Kupang.

Penyelesaian penyusunan Tugas Akhir ini tidak terlepas dari bantuan beberapa pihak. Oleh karena itu, pada kesempatan ini, ingin diucapkan terima kasih kepada:

1. Pater Dr. Stefanus Lio, SVD, S.Fil., M.A., selaku Rektor Universitas Katolik Widya Mandira Kupang.
2. Bapak Dr. Don Gaspar N. Da Costa, S.T., M.T., selaku Dekan Fakultas Teknik Universitas Katolik Widya Mandira.
3. Ibu Yovinia Carmeneja Hoar Siki, S.T., M.T., selaku Ketua Program Studi Ilmu Komputer Universitas Katolik Widya Mandira.
4. Kedua dosen pembimbing, Ibu Yulianti Paula Bria, S.T., M.T., Ph.D., selaku Pembimbing I dan Bapak Donatus Joseph Manehat, S.Si., M.Kom., selaku Pembimbing II, atas bimbingan, arahan, kritik, dan saran yang diberikan selama proses penyusunan Tugas Akhir ini. Tanpa bimbingan yang tulus dan penuh kesabaran, Tugas Akhir ini tidak mungkin dapat terselesaikan.
5. Kedua dosen penguji, Ibu Sisilia Daeng Bakka Mau, S.Kom., M.T.,

selaku Penguji I dan Bapak Paul Filson Mite Tengangatu, S.Kom. M.T.I.,
selaku Penguji II.

6. Ibu Yovina Carmeneja Hoar Siki, S.T., M.T., sebagai dosen Pembimbing Akademik yang telah memberikan arahan dan dukungan dari awal semester hingga penyelesaian Tugas Akhir ini.
7. Bapak dan Ibu Pegawai Tata Usaha dan Laboran yang telah membantu selama ini.
8. Seluruh Dosen Program Studi Ilmu Komputer di Universitas Katolik Widya Mandira Kupang.

Penyusunan Tugas Akhir ini tidak luput dari kesalahan. Untuk itu, diharapkan saran dan kritik dari berbagai pihak sehingga Tugas Akhir ini dapat memberikan manfaat bagi semua pihak.

Kupang, Januari 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSEMBAHAN.....	iv
MOTTO	v
PERNYATAAN DAN KEASLIAN HASIL KARYA.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR GRAFIK	xvi
ABSTRAK	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	4
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan.....	6
1.7 Daftar Istilah dan Singkatan.....	8
BAB II LANDASAN TEORI	13

2.1	Penelitian Terdahulu.....	13
2.2	Landasan Teori	27
2.2.1	<i>Cyber Attack</i>	27
2.2.2	<i>Cybersecurity</i>	29
2.2.3	<i>Firewall & Web Application Firewall (WAF)</i>	30
2.2.4	<i>Machine Learning</i>	32
2.2.5	<i>Pre-processing Data</i>	32
2.2.6	<i>Cleaning Data</i>	33
2.2.7	<i>Feature Engineering</i>	34
2.2.8	<i>Transformation & Scaling</i>	34
2.2.9	<i>Feature Selection & Feature Importance</i>	35
2.2.10	<i>Handling Imbalance Data (SMOTE)</i>	36
2.2.11	<i>Cross Validation</i>	36
2.2.12	<i>Tree-based Classifier Model</i>	37
2.2.13	<i>Hyperparameter Tuning</i>	44
2.2.14	Evaluasi Kinerja Model.....	44
2.2.15	<i>Jupyter Notebook</i>	47
BAB III	METODOLOGI PENELITIAN	48
3.1	Karakteristik Data	48
3.2	<i>Pre-processing Data</i>	49

3.2.1	<i>Cleaning Data</i>	49
3.2.2	<i>Feature Engineering</i>	50
3.2.3	<i>Transformation & Scaling</i>	51
3.2.4	<i>Feature Selection (Random Forest Feature Importance)</i>	52
3.2.5	<i>Handling Imbalance Data (SMOTE)</i>	53
3.3	Pemodelan.....	54
3.4	<i>Hyperparameter Tuning</i>	60
3.5	Evaluasi Kinerja Model.....	60
3.6	Perbandingan Model	61
3.7	Definisi Operasional.....	61
BAB IV HASIL DAN PEMBAHASAN		68
4.1	Karakteristik Data	68
4.2	<i>Pre-processing Data</i>	71
4.2.1	<i>Cleaning Data</i>	75
4.2.2	<i>Feature Engineering</i>	78
4.2.3	<i>Transformation & Scaling</i>	80
4.2.4	<i>Feature Selection (Random Forest Feature Importance)</i>	81
4.2.5	<i>Handling Imbalance Data (SMOTE)</i>	83
4.3	Pemodelan & Evaluasi Model.....	84
BAB V PENGUJIAN DAN ANALISIS HASIL		94

BAB VI PENUTUP	102
6.1 Kesimpulan	102
6.2 Saran.....	104
DAFTAR PUSTAKA.....	105
SURAT KETERANGAN HASIL CEK PLAGIASI.....	110

DAFTAR TABEL

Tabel 1.1. Daftar Istilah dan Singkatan.....	8
Tabel 2.1. Daftar Penelitian Terdahulu.....	17
Tabel 2.2. <i>Confusion Matrix</i>	45
Tabel 3.1. Distribusi Data.....	54
Tabel 3.2. Ilustrasi <i>Dataset</i> Deteksi Serangan Sederhana.....	55
Tabel 3.3. Definisi Operasional.....	61
Tabel 4.1. Daftar Atribut dan Deskripsi CSIC 2010 HTTP <i>Dataset</i>	69
Tabel 4.2. Alasan Penghapusan Atribut.....	71
Tabel 4.3. Informasi Data.....	73
Tabel 4.4. Hasil Pengecekan Nilai Atribut Kategorikal	74
Tabel 4.5. Hasil Pemodelan.....	85
Tabel 5.1. Hasil <i>Hyperparameter Tuning</i>	96
Tabel 5.2. Hasil Persentase <i>Hyperparameter Tuning</i>	96

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi <i>Decision Tree</i>	38
Gambar 2.2. Ilustrasi <i>Random Forest</i>	40
Gambar 2.3. Ilustrasi <i>Gradient Boosting Machine</i>	41
Gambar 2.4. Ilustrasi <i>XGBoost</i>	43
Gambar 3.1. Tahapan Penelitian	48
Gambar 3.2. <i>Decision Tree</i>	56
Gambar 3.3. <i>Random Forest</i>	57
Gambar 4.1. <i>Source Code</i> Konfigurasi & Muat <i>Dataset</i>	71
Gambar 4.2. <i>Source Code</i> Hapus Atribut Tidak Relevan.....	76
Gambar 4.3. <i>Source Code</i> Normalisasi <i>length</i>	76
Gambar 4.4. <i>Source Code</i> Handling <i>Missing Values</i>	77
Gambar 4.5. <i>Source Code</i> Normalisasi URL & <i>content</i>	77
Gambar 4.6. <i>Source Code</i> Hapus Duplikasi Data	78
Gambar 4.7. <i>Output Cleaning Data</i>	78
Gambar 4.8. <i>Source Code</i> Hitung Entropi URL dan <i>content</i>	79
Gambar 4.9. <i>Source Code</i> Ekstrasi Panjang URL, <i>content</i> , & Karakter Spesial ..	79
Gambar 4.10. <i>Source Code</i> Transformasi <i>Method</i>	80
Gambar 4.11. <i>Source Code</i> Hapus Fitur Awal & Pisah <i>classification</i>	80
Gambar 4.12. <i>Output Feature Engineering</i>	80
Gambar 4.13. <i>Source Code</i> <i>MinMaxScaler</i>	81
Gambar 4.14. <i>Source Code</i> <i>Feature Selection - Random Forest</i> FI.....	82
Gambar 4.15. Daftar 10 Fitur Terpilih	83

Gambar 4.16. <i>Handling Imbalance Data (SMOTE)</i>	84
Gambar 4.17. <i>Output Handling Imbalance Data - SMOTE</i>	84
Gambar 4.18. <i>Source Code Pelatihan & Evaluasi Model</i>	86
Gambar 4.19. <i>Source Code Confusion Matrix Pemodelan</i>	87
Gambar 4.20. <i>Output Confusion Matrix Pemodelan</i>	87
Gambar 5.1. <i>Source Code Hyperparameter Tuning</i>	95
Gambar 5.2. <i>Output Confusion Matrix Hyperparameter Tuning</i>	97
Gambar 5.3. <i>Source Code Visualisasi Grafik</i>	97
Gambar 5.4. <i>Output Visualisasi Grafik</i>	98

DAFTAR GRAFIK

Grafik 4.1. Performa Model (<i>K-Fold</i> Terbaik).....	91
Grafik 4.2. Waktu Komputasi per <i>K-Fold</i>	92
Grafik 5.1. Performa Model (<i>K-Fold</i> Terbaik).....	98
Grafik 5.2. Waktu Komputasi per <i>K-Fold</i>	99

ABSTRAK

Aplikasi web merupakan fondasi utama operasional di berbagai sektor strategis, namun kerentanannya terhadap serangan siber menjadi ancaman serius bagi keamanan data. Mekanisme pertahanan konvensional sering kali kesulitan menghadapi variasi serangan baru yang semakin kompleks. Penelitian terdahulu banyak menggunakan pendekatan *Natural Language Processing* (NLP) yang memiliki kompleksitas komputasi tinggi sehingga kurang optimal untuk implementasi *real-time*. Tujuan penelitian ini adalah untuk menentukan fitur paling signifikan menggunakan *Feature Importance* dengan algoritma *Random Forest* serta membandingkan kinerja model *Tree-Based Classifier* (*Decision Tree*, *Random Forest*, *GBM*, dan *XGBoost*) dalam mengklasifikasikan lalu lintas HTTP menggunakan dataset CSIC 2010. Hasil evaluasi menggunakan *Stratified K-Fold Cross Validation* dan *Hyperparameter Tuning* menunjukkan bahwa model *Random Forest* merupakan model yang paling unggul secara keseluruhan dengan nilai *F1-score* 92,71% dan *Recall* 94,95% pada *K-Fold* = 7, meskipun membutuhkan waktu komputasi 289,76 detik. Di sisi lain, *XGBoost* menawarkan kompromi terbaik dengan *F1-score* 92% dan waktu eksekusi hanya 39,45 detik. Sementara itu, *Decision Tree* menjadi alternatif paling efisien untuk sistem yang mengutamakan kecepatan dengan waktu komputasi hanya 2,05 detik. Penelitian yang dilakukan membuktikan bahwa pendekatan *Tree-Based Classifier* non-NLP mampu memberikan performa deteksi yang kompetitif, lebih ringan, dan memiliki interpretabilitas tinggi, sehingga sangat optimal untuk implementasi pada sistem *Web Application Firewall* modern.

Kata Kunci: *Web Application Firewall, Machine Learning, Tree-Based Classifier, CSIC 2010, Feature Importance.*

ABSTRACT

Web applications serve as the main operational foundation of various strategic sectors, yet their vulnerability to cyberattacks poses a severe threat to data security. Conventional defense mechanisms often struggle to mitigate increasingly complex new attack variations. Previous research has extensively utilized Natural Language Processing (NLP) approaches, which entail high computational complexity, making them less optimal for real-time implementation. The objective of this research is to identify the most significant features using the Feature Importance with Random Forest algorithm and to compare the performance of Tree-Based Classifiers (Decision Tree, Random Forest, GBM, and XGBoost) in classifying HTTP traffic using CSIC 2010 dataset. Evaluation using Stratified K-Fold Cross Validation and Hyperparameter Tuning demonstrated that the Random Forest model was the most superior overall, achieving an F1-score of 92.71% and a Recall of 94.95% at K-Fold = 7, despite requiring a computation time of 289.76 seconds. Conversely, XGBoost offered the best compromise with an F1-score of 92% and an execution time of only 39.45 seconds. Meanwhile, the Decision Tree model emerged as the most efficient alternative for systems prioritizing speed, with a computation time of just 2.05 seconds. Research that held proves that the non-NLP Tree-Based Classifier approach provides competitive detection performance, is lightweight, and possesses high interpretability, making it highly optimal for modern Web Application Firewall implementations.

Keywords: *Web Application Firewall, Machine Learning, Tree-Based Classifier, CSIC 2010, Feature Importance.*