

**DETEKSI WEBSITE PHISHING BERBASIS
KARAKTERISTIK URL MENGGUNAKAN
PENDEKATAN RANDOM FOREST DAN
DECISION TREE**



SKRIPSI

**LOUIS GAUDIAMO HALIM
00000062782**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**DETEKSI WEBSITE PHISHING BERBASIS
KARAKTERISTIK URL MENGGUNAKAN
PENDEKATAN RANDOM FOREST DAN
DECISION TREE**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

LOUIS GAUDIAMO HALIM
00000062782

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Louis Gaudiamo Halim
Nomor Induk Mahasiswa : 00000062782
Program Studi : Informatika

Skripsi dengan judul:

Deteksi Website Phishing Berbasis Karakteristik URL Menggunakan Pendekatan Random Forest dan Decision Tree

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 08 Januari 2026



(Louis Gaudiamo Halim)

HALAMAN PENGESAHAN

Skripsi dengan judul

DETEKSI WEBSITE PHISHING BERBASIS KARAKTERISTIK URL MENGUNAKAN PENDEKATAN RANDOM FOREST DAN DECISION TREE

oleh

Nama : Louis Gaudiamo Halim
NIM : 00000062782
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Kamis, 15 Januari 2026

Pukul 15.00 s/s 17.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

(Wirawan Istiono, S.Kom., M.Kom)

NIDN: 0313048304

Penguji

(Fenina Adline Twince Tobing, S.Kom.,

M.Kom)

NIDN: 0406058802

Pembimbing

(SY.Yuliani, S.Kom., MT., Ph.D., CEH)

NIDN: 0411037904

Ketua Program Studi Informatika,

(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)

NIDN: 0315109103

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Louis Gaudiamo Halim
NIM : 00000062782
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Deteksi Website Phishing Berbasis
Karakteristik URL Menggunakan
Pendekatan Random Forest dan
Decision Tree

Menyatakan dengan sesungguhnya bahwa saya bersedia (**pilih salah satu**):

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
- Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

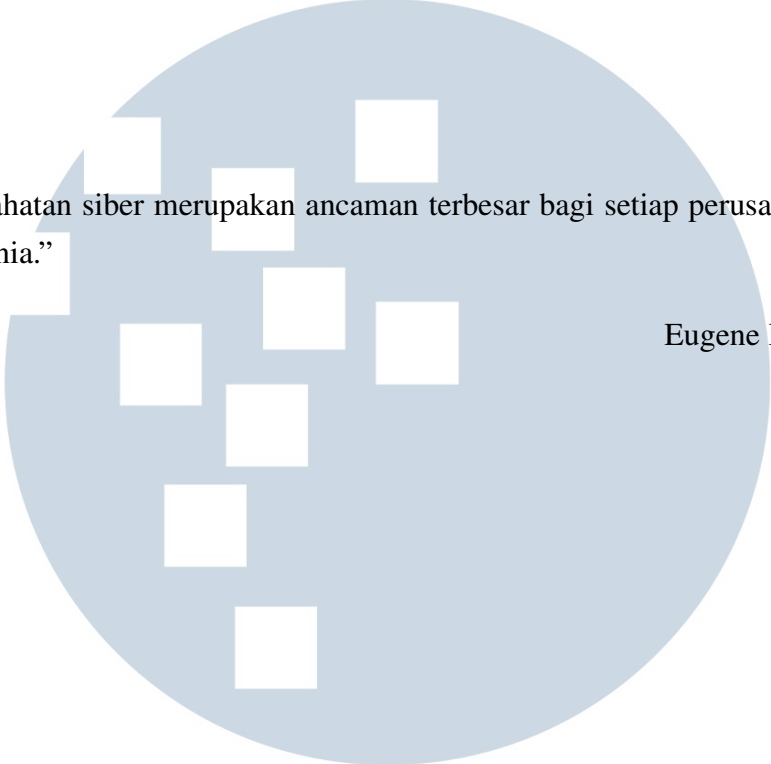
Tangerang, 08 Januari 2026

Yang menyatakan



Louis Gaudiamo Halim

HALAMAN PERSEMBAHAN / MOTTO



”Kejahatan siber merupakan ancaman terbesar bagi setiap perusahaan di dunia.”

Eugene Kaspersky

UMN

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan karunia-Nya, sehingga laporan Tugas Akhir dengan judul "Deteksi Website Phishing Berbasis Karakteristik URL Menggunakan Pendekatan Random Forest dan Decision Tree" dapat terselesaikan dengan baik. Laporan ini disusun untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer di Universitas Multimedia Nusantara. Penulis menyadari bahwa kelancaran penyusunan laporan ini tidak lepas dari bimbingan dan dukungan berbagai pihak. Mengucapkan terima kasih

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu SY.Yuliani.,S.Kom.,MT.,Ph.D.,CEH, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
6. Teman-teman yang telah membantu dan mendukung saya selama penulisan laporan skripsi ini.

Semoga laporan Tugas Akhir ini dapat memberikan manfaat dan kontribusi positif bagi pembaca serta pengembangan ilmu pengetahuan.

Tangerang, 08 Januari 2026



Louis Gaudiamo Halim

DETEKSI WEBSITE PHISHING BERBASIS KARAKTERISTIK URL MENGUNAKAN PENDEKATAN RANDOM FOREST DAN DECISION TREE

Louis Gaudiamo Halim

ABSTRAK

Serangan *phishing* merupakan teknik rekayasa sosial yang menjadi salah satu ancaman keamanan siber terbesar yang bertujuan mencuri informasi sensitif dengan memalsukan situs web agar terlihat asli. Tantangan utama sistem deteksi saat ini adalah bergantung pada analisis isi halaman (*content-based*) sering kali membebani sistem karena membutuhkan daya komputasi tinggi dan waktu proses yang lama. Untuk mengatasi kendala tersebut, penelitian ini menawarkan solusi yang lebih ringan dan cepat, yaitu sistem pendeteksi *phishing* hanya dengan menganalisis karakteristik URL (*lexical-based features*) dan reputasi domain, tanpa perlu mengunduh konten halaman. Dengan memanfaatkan dataset publik dari Kaggle yang mencakup 11.430 sampel URL yang terdiri dari 89 kolom fitur, studi ini menguji ketangguhan algoritma *Random Forest* melawan model dasar *Decision Tree*. Melalui serangkaian pengujian fitur leksikal dan optimasi *hyperparameter tuning* untuk hasil maksimal. Hasil eksperimen membuktikan keunggulan *Random Forest*, yang berhasil mencatat akurasi (*accuracy*) hingga 95.51%, dengan presisi (*precision*) 95.51%, *recall* 95.43%, dan *F1-score* 95.47%, hasil yang lebih stabil dibandingkan *Decision Tree* di angka 92.52%. Temuan ini menegaskan bahwa kombinasi fitur sederhana seperti umur domain dan status indeks pencarian sudah cukup untuk membangun pertahanan anti-*phishing* yang cepat dan akurat.

Kata kunci: Karakteristik URL, Keamanan Siber, *Machine Learning*, *Phishing*, *Random Forest*

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

PHISHING WEBSITE DETECTION BASED ON URL FEATURES USING RANDOM FOREST AND DECISION TREE

Louis Gaudiamo Halim

ABSTRACT

Phishing attacks are social engineering techniques representing one of the greatest cybersecurity threats, aiming to steal sensitive information by falsifying websites to appear authentic. The main challenge of current detection systems lies in their reliance on content-based analysis, which often burdens the system due to high computational power requirements and long processing times. To address these constraints, this study offers a lighter and faster solution include a phishing detection system that analyzes only URL characteristics (lexical-based features) and domain reputation, eliminating the need to download page content. Utilizing a public dataset from Kaggle comprising 11,430 URL samples and 89 feature columns, this study evaluates the robustness of the Random Forest algorithm against a baseline Decision Tree model. The process involves a series of lexical feature tests and hyperparameter tuning optimization to achieve maximum results. Experimental results demonstrate the superiority of Random Forest, achieving an accuracy of 95.51%, with a precision of 95.51%, recall of 95.43%, and F1-score of 95.47%, reflecting significantly higher stability compared to the Decision Tree at 92.52%. These findings confirm that a combination of simple features, such as domain age and search index status, is sufficient to construct a fast and accurate anti-phishing defense.

Keywords: *Cybersecurity, Machine Learning, Phishing, Random Forest, URL Characteristics*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR RUMUS	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	6
1.3 Batasan Permasalahan	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
1.6 Sistematika Penulisan	8
BAB 2 LANDASAN TEORI	10
2.1 Konsep Dasar Website	10
2.2 HyperText Transfer Protocol (HTTP/HTTPS)	11
2.3 Teori Dasar Cyber Security	12
2.4 Metode CIA Triad	14
2.5 Cyber Security Threat	15
2.5.1 Macam-macam Cyber Threat	16
2.6 Phishing	21
2.6.1 Struktur URL Phishing	27
2.7 Machine Learning	30
2.7.1 Supervised Learning	31
2.7.2 Ensemble Learning	32
2.7.3 Klasifikasi	33
2.7.4 Confusion Matrix	33
2.8 Data Mining	34
2.8.1 CRISP-DM	35
2.9 Algoritma	37
2.9.1 Decision Tree	38
2.9.2 Classification and Regression Trees (CART)	38
2.9.3 Random Forest (RF)	39
2.10 Tools	40
2.10.1 Python	40
2.10.2 Kaggle	41
2.10.3 Google Colab	42
2.11 Penelitian Terdahulu	43
2.11.1 Aljofey et al. (2022)	44
2.11.2 Yang et al. (2021)	45
2.11.3 Ryan Putra Ramadhan dan Teti Desyani (2023)	46

2.11.4	Fatiha et al. (2024)	47
2.11.5	Bari et al. (2025)	48
2.11.6	Alsariera et al. (2020)	49
BAB 3	METODOLOGI PENELITIAN	52
3.1	Gambaran Umum Objek Penelitian	52
3.2	Metode Penelitian	53
3.2.1	Tahapan Penelitian	53
3.2.2	Pendekatan dan Jenis Penelitian	55
3.3	Data Penelitian	56
3.3.1	Dataset	56
3.3.2	Teknik Pengumpulan Data	57
3.4	Variabel Penelitian	57
3.4.1	Variabel Independen (X)	58
3.4.2	Variabel Dependen (Y)	61
3.5	Proses Analisis dan Pembangunan Model	61
3.5.1	Data Collection	62
3.5.2	Tahapan Pengolahan Data (Data Preprocessing)	63
3.5.3	Penerapan Model Klasifikasi (Decision Tree dan Random Forest)	66
3.5.4	Evaluasi Model	67
3.5.5	Proses Pembentukan Pohon Keputusan	71
3.5.6	Ekstraksi Fitur (Feature Extraction)	74
BAB 4	HASIL DAN DISKUSI	75
4.1	Tahapan Persiapan Lingkungan Eksperimen	75
4.1.1	Konfigurasi Tools dan Libraries	75
4.1.2	Load Dataset	77
4.2	Hasil Pra-pemrosesan Data (Data Preprocessing)	78
4.2.1	Seleksi Fitur (<i>Feature Selection</i>)	79
4.2.2	Pembersihan Data (<i>Data Cleaning</i>)	82
4.2.3	Transformasi Data	85
4.2.4	Pembagian Data (<i>Data Splitting</i>)	85
4.3	Implementasi dan Evaluasi Model Decision Tree	86
4.3.1	Arsitektur Pipeline dan Konfigurasi Model	87
4.3.2	Pelatihan Model Decision Tree	87
4.3.3	Analisis Overfitting Model Decision Tree	88
4.3.4	Evaluasi Confusion Matrix Decision Tree	89
4.3.5	Visualisasi Struktur Pohon Decision Tree	91
4.4	Implementasi dan Evaluasi Model Random Forest	94
4.4.1	Konfigurasi Model Random Forest	94
4.4.2	Hyperparameter Tuning Random Forest	94
4.4.3	Evaluasi Model Random Forest	96
4.4.4	Confusion Matrix Random Forest	97
4.4.5	Analisis ROC Curve dan AUC	99
4.4.6	Cross Validation Random Forest	100
4.5	Visualisasi Features Importance	101
4.5.1	Feature Extraction	103
4.5.2	Implementasi Sistem Deteksi Karakteristik URL	107
4.6	Perbandingan Model Decision Tree dan Random Forest	110
BAB 5	SIMPULAN DAN SARAN	112
5.1	Simpulan	112
5.2	Saran	113

DAFTAR PUSTAKA	115
--------------------------	-----



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR TABEL

Tabel 2.1	Ringkasan Penelitian Terdahulu Deteksi Website Phishing	51
Tabel 3.1	Variabel Independen Berdasarkan Karakteristik URL (Bagian 1)	59
Tabel 3.2	Variabel Independen Berdasarkan Karakteristik URL (Bagian 2)	60
Tabel 3.3	Keterangan Variabel Dependen (Label Keluaran)	61
Tabel 3.4	Tabel Konsep <i>Confusion Matrix</i> pada Model Klasifikasi . .	69
Tabel 4.1	Konfigurasi Pustaka (<i>Libraries</i>) dan Modul Python yang Digunakan	76
Tabel 4.2	Daftar Fitur Terpilih untuk Deteksi URL Phishing (Bagian 1)	80
Tabel 4.3	Lanjutan Daftar Fitur Terpilih (Bagian 2)	81
Tabel 4.4	Parameter Terbaik Hasil <i>Hyperparameter Tuning Random Forest</i>	96
Tabel 4.5	Confusion Matrix Model Random Forest (Tuned)	99
Tabel 4.6	Perbandingan Kinerja Model Decision Tree dan Random Forest	111



DAFTAR GAMBAR

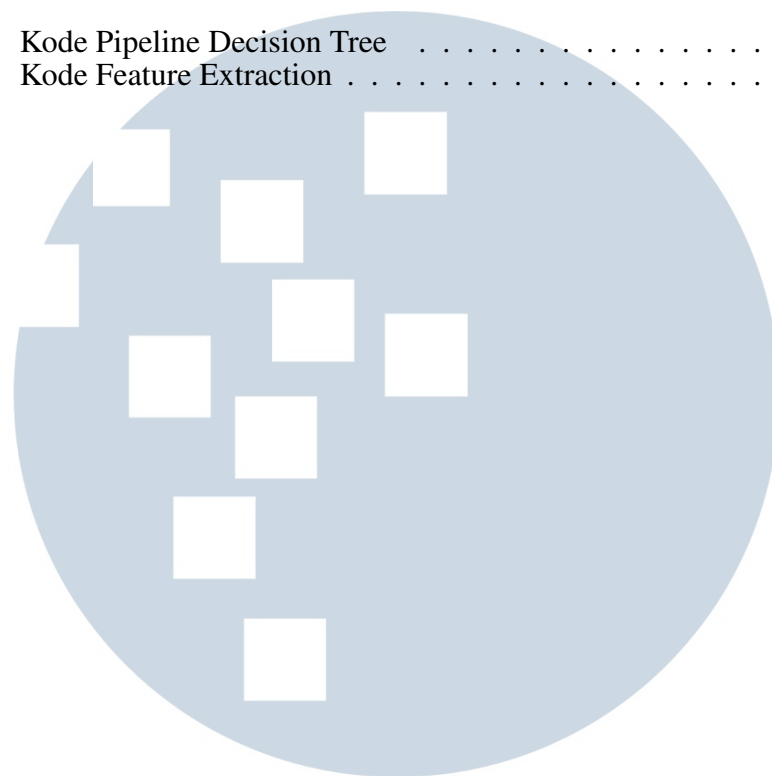
Gambar 1.1	Diagram Batang Kasus <i>Phishing</i> di Indonesia pada kuartal pertama.	2
Gambar 1.2	Diagram Batang Kasus <i>Phishing</i> di Indonesia pada kuartal kedua.	3
Gambar 1.3	Diagram <i>Pie Chart</i> Target Industri Serangan <i>Phishing</i> . . .	5
Gambar 2.1	Perbedaan Protokol HTTP dan HTTPS.	11
Gambar 2.2	Kutipan Kevin David Mitnick.	13
Gambar 2.3	Elemen CIA Triad.	14
Gambar 2.4	Jenis-jenis Ancaman Malware.	16
Gambar 2.5	Jenis-jenis Insider Threat.	17
Gambar 2.6	Serangan DoS dilakukan "sendirian".	18
Gambar 2.7	Serangan DDoS dilakukan oleh "grup".	18
Gambar 2.8	Ancaman MITM.	19
Gambar 2.9	Passive Capture Device.	20
Gambar 2.10	Gambar Detail dari Rekayasa Sosial atau <i>Social Engineering</i>	21
Gambar 2.11	Contoh Penipuan WhatsApp.	22
Gambar 2.12	Contoh <i>Email Phishing</i>	24
Gambar 2.13	Proses <i>Spear Phishing</i>	25
Gambar 2.14	Perbedaan <i>Phishing</i> Reguler dan <i>Phishing</i> Whaling.	26
Gambar 2.15	Contoh Kemiripan Antara Website <i>Phishing</i> dan Website Asli Suatu Instansi.	27
Gambar 2.16	Struktur Panjang URL.	28
Gambar 2.17	Contoh Typosquatting dan Homograph Attack.	30
Gambar 2.18	<i>Supervised Learning</i>	32
Gambar 2.19	Confusion Matrix.	34
Gambar 2.20	Diagram CRISP-DM.	35
Gambar 2.21	Decision Tree	38
Gambar 2.22	Pohon Keputusan Random Forest.	40
Gambar 2.23	Logo Bahasa Pemrograman Python	41
Gambar 2.24	Platform Dataset Kaggle.	42
Gambar 2.25	Logo Platform Google Colab.	43
Gambar 2.26	Framework Penelitian Aljofey et al. (2022).	45
Gambar 2.27	Framework Penelitian Yang et al. (2021).	46
Gambar 2.28	Framework Penelitian Fatiha et al. (2024).	48
Gambar 2.29	Framework Penelitian Bari et al. (2025).	49
Gambar 2.30	Framework Penelitian Alsariera et al. (2020).	50
Gambar 3.1	Tahapan Penelitian.	54
Gambar 3.2	Dataset Kaggle.	56
Gambar 3.3	Tahapan Data Collection.	62
Gambar 3.4	Tahapan <i>Data Preprocessing</i>	65
Gambar 3.5	Tahapan <i>Modeling</i>	67
Gambar 3.6	Tahapan <i>Evaluation</i>	68
Gambar 3.7	Proses Pembentukan Pohon Keputusan.	73
Gambar 4.1	Output Menampilkan Tabel Data.	78
Gambar 4.2	Output Feature Selection.	82
Gambar 4.3	Output Pengecekan Nilai <i>Missing Value</i> dan Duplikat.	83
Gambar 4.4	Output Penghapusan Nilai Duplikat.	83

Gambar 4.5	Output Perbandingan <i>Phishing</i> dan <i>Legitimate</i>	84
Gambar 4.6	Output Label Encoding.	85
Gambar 4.7	Output Pembagian Data Latih dan Uji.	86
Gambar 4.8	Output Pelatihan Model Decision Tree.	88
Gambar 4.9	Output Overfitting Model Decision Tree.	89
Gambar 4.10	Output <i>Confusion Matrix</i> Model Decision Tree	90
Gambar 4.11	Visualisasi Struktur Pohon Decision Tree (<i>Top 3 Levels</i>). . .	91
Gambar 4.12	Output <i>Hyperparameter Tuning</i> Random Forest.	95
Gambar 4.13	Hasil Evaluasi Model Final <i>Random Forest</i>	97
Gambar 4.14	<i>Confusion Matrix</i> Model <i>Random Forest</i>	98
Gambar 4.15	Grafik ROC dan AUC.	100
Gambar 4.16	Visualisasi Riwayat Akurasi pada 50 Kombinasi <i>Hyperparameter</i>	101
Gambar 4.17	Diagram Batang <i>Features Importance</i>	102
Gambar 4.18	Hasil Testing URL https://www.umn.ac.id	108
Gambar 4.19	Hasil Testing Multiple URL's.	109



DAFTAR KODE

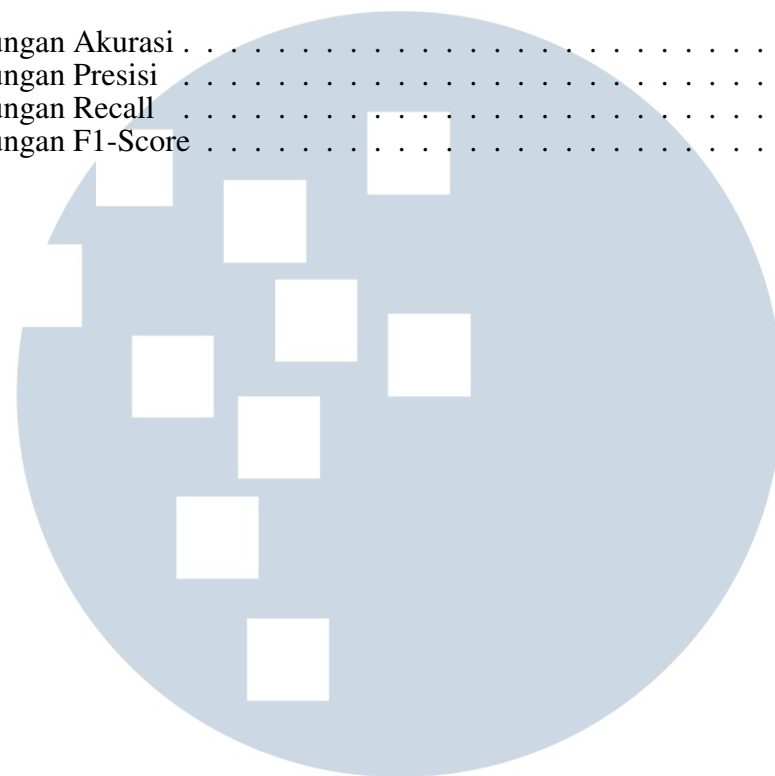
Kode 4.1	Kode Pipeline Decision Tree	87
Kode 4.2	Kode Feature Extraction	104



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR RUMUS

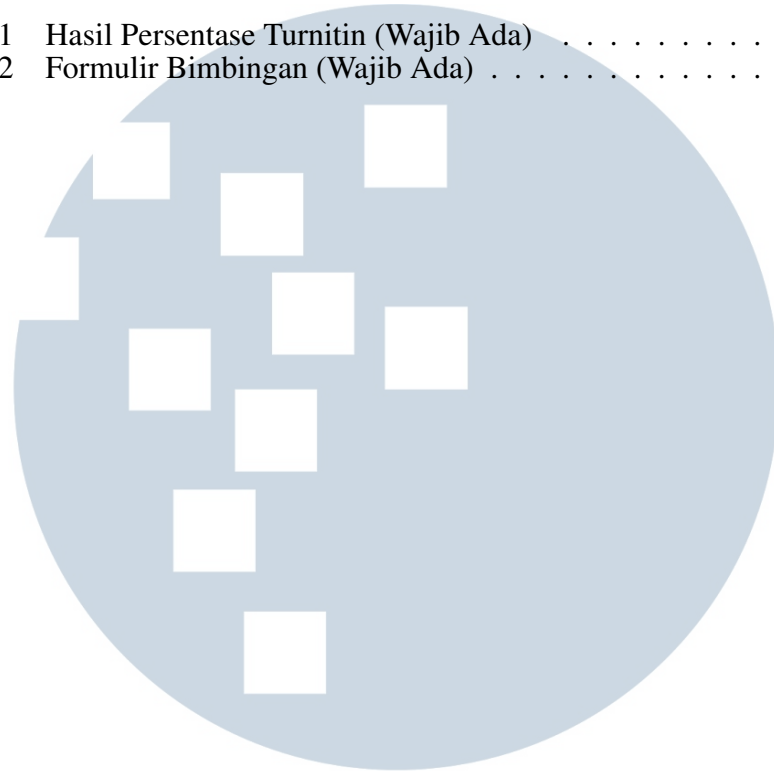
3.1 Perhitungan Akurasi	69
3.2 Perhitungan Presisi	70
3.3 Perhitungan Recall	70
3.4 Perhitungan F1-Score	70



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin (Wajib Ada)	121
Lampiran 2	Formulir Bimbingan (Wajib Ada)	142



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA