

**DETEKSI SERANGAN *PHISHING* BERBASIS URL
DENGAN *VARIATIONAL AUTOENCODER* DAN
CONVOLUTIONAL NEURAL NETWORK
(VAE-CNN)**



Skripsi

Untuk memenuhi syarat memperoleh Derajat

Sarjana Teknik (S.T.)

Oleh:

YOGA SYAHPUTRA

NIM 2101020105

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK ELEKTRO DAN INFORMATIKA
FAKULTAS TEKNIK DAN TEKNOLOGI KEMARITIMAN
UNIVERSITAS MARITIM RAJA ALI HAJI
TANJUNGPINANG
2025**

**DETEKSI SERANGAN *PHISHING* BERBASIS URL
DENGAN *VARIATIONAL AUTOENCODER* DAN
CONVOLUTIONAL NEURAL NETWORK
(VAE-CNN)**



Skripsi

Untuk memenuhi syarat memperoleh Derajat

Sarjana Teknik (S.T.)

Oleh:

YOGA SYAHPUTRA

NIM 2101020105

Telah mengetahui dan disetujui oleh:

Pembimbing I,

A handwritten signature in blue ink, likely belonging to Hendra Kurniawan, the first supervisor.

Hendra Kurniawan, S.Kom., M.Sc.Eng., Ph.D.
NIP. 198404022014041001

Pembimbing II,

A handwritten signature in blue ink, likely belonging to Novrizal Fattah Fahmitra, the second supervisor.

Novrizal Fattah Fahmitra, S.Kom., M.Kom.
NIP. 198911012024061002

HALAMAN PENGESAHAN

Judul Skripsi : Deteksi Serangan *Phishing* Berbasis URL dengan
Variational Autoencoder dan *Convolutional Neural*
Network (VAE-CNN)

Nama Mahasiswa : Yoga Syahputra

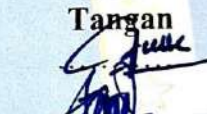
NIM : 2101020105

Jurusan : Teknik Elektro dan Informatika

Program Studi : Teknik Informatika

Telah dipertahankan di depan Dewan Penguji dan dinyatakan lulus
pada tanggal 5 Juni 2025

Susunan Tim Pembimbing dan Penguji

Jabatan	Nama Dosen	Tanda Tangan	Tanggal
Pembimbing I	: Hendra Kurniawan, S.Kom., M.Sc.Eng., Ph.D.		17/6-25
Pembimbing II	: Novrizal Fattah Fahmitra, S.Kom., M.Kom.		20/6-25
Ketua Penguji	: Ferdi Chahyadi, S.Kom., M.Cs.		16/6-25
Anggota Penguji I	: Rifaldi Herikson, S.Kom., M.Kom.		18/6 25
Anggota Penguji II	: Berta Erwin SLAM, S.T., M.Kom.		18/6

Tanjungpinang, 1 Juli 2025

Universitas Maritim Raja Ali Haji

Fakultas Teknik dan Teknologi Kemaritiman

Dekan Fakultas Teknik dan Teknologi Kemaritiman



Martalehi Bettiza, S.Si., M.Sc.

NIPPPK: 197508282021212006

PERNYATAAN ORISINALITAS

Dengan ini menyatakan bahwa skripsi saya yang berjudul “Deteksi Serangan *Phishing* berbasis URL dengan *Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN)” adalah benar karya saya dengan arahan dari komisi pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Jika kemudian hari ternyata terbukti pernyataan saya ini tidak benar dan melanggar peraturan yang sah dalam karya tulis dan hak intelektual maka saya bersedia ijazah yang telah saya terima untuk ditarik kembali oleh Universitas Maritim Raja Ali Haji.

Tanjungpinang, 30 Juni 2025



(Yoga Syahputra)

KATA PENGANTAR

Puji syukur ke hadirat Allah *Subhanahu wa Ta'ala* atas limpahan rahmat, karunia, dan hidayah-Nya yang tak terkira sehingga penulis dapat menyelesaikan skripsi yang berjudul “Deteksi Serangan *Phishing* berbasis URL dengan *Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN)”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik informatika, Jurusan Teknik Elektro dan Informatika, Fakultas Teknik dan Teknologi Kemaritiman, Universitas Maritim Raja Ali Haji (UMRAH).

Selanjutnya, penulis ingin menyampaikan ribuan terima kasih dari lubuk hati yang terdalam kepada berbagai pihak, termasuk kedua orang tua, saudara-saudari, keluarga besar, para dosen, dan sahabat-sahabat tercinta atas doa, dukungan, bimbingan, dan bantuan yang telah diberikan selama proses penyusunan skripsi. Tanpa kalian, penulis tidak akan bisa berada di titik ini. Semoga Allah *Subhanahu wa Ta'ala*, Tuhan Yang Maha Esa membalas semua kebaikan yang telah diberikan kepada penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, sehingga kritik yang bersifat konstruktif sangat diharapkan untuk yang lebih baik lagi ke depannya. Semoga skripsi ini dapat bermanfaat dan tentunya berkontribusi terhadap pengembangan ilmu pengetahuan di masa mendatang, terutama dalam bidang keamanan siber berbasis kecerdasan buatan, *aamiin yaa rabbal alamin*.

Tanjungpinang, 30 Juni 2025



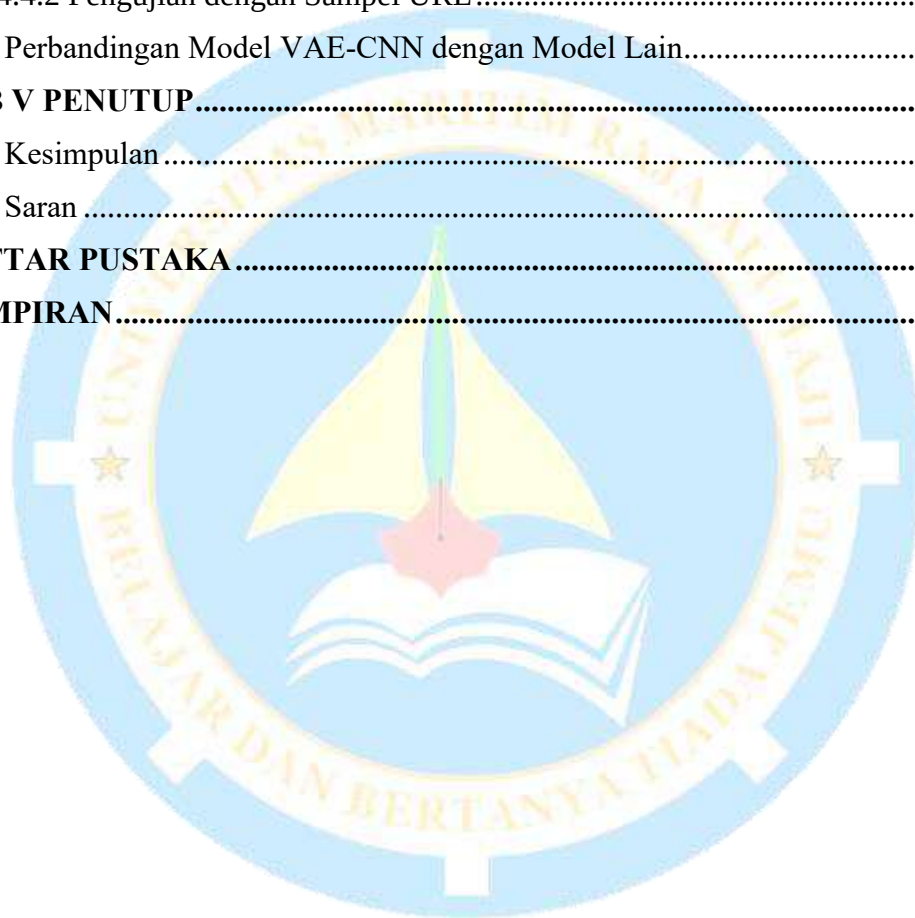
(Yoga Syahputra)

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
PERNYATAAN ORISINALITAS.....	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
GLOSARIUM.....	xv
ABSTRAK	xvi
<i>ABSTRACT</i>	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah	6
1.3 Batasan Masalah	6
1.4 Tujuan Penelitian	7
1.5 Manfaat Penelitian	7
1.6 Sistematika Penulisan	8
BAB II KAJIAN LITERATUR	9
2.1 Kajian Literatur.....	9
2.2 Landasan Teori	20
2.2.1 <i>Phishing</i>	20
2.2.1.1 Konsep <i>Phishing</i>	20
2.2.1.2 Metode Deteksi <i>Phishing</i>	20
2.2.2 <i>Uniform Resource Locator (URL)</i>	22
2.2.3 <i>Deep Learning</i>	23
2.2.4 Deteksi Anomali (<i>Anomaly Detection</i>).....	24
2.2.5 Python	24

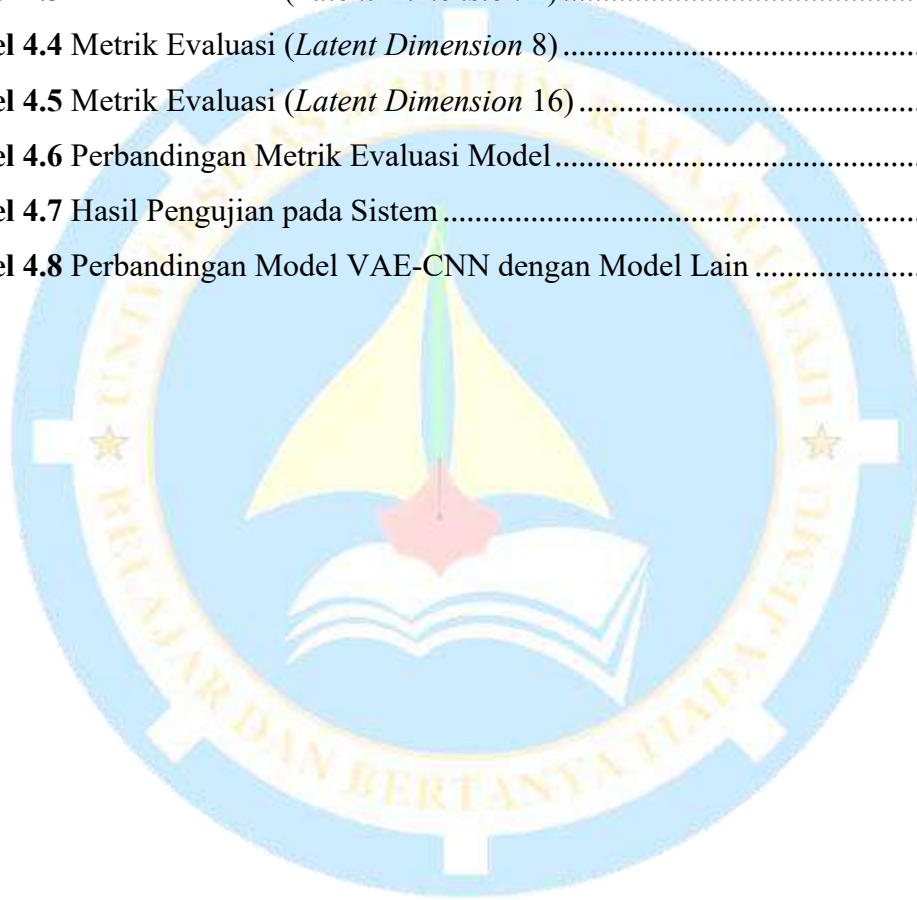
2.2.6 Variational Autoencoder (VAE).....	25
2.2.7 Convolutional Neural Network (CNN).....	26
2.2.8 Variational Autoencoder & Convolutional Neural Network (VAE-CNN).....	27
2.2.9 Performa Model	32
2.2.9.1 Seleksi <i>Threshold</i> untuk Deteksi Anomali: Metode Persentil	32
2.2.9.2 Metrik Evaluasi	33
BAB III METODE PENELITIAN	35
3.1 Sub Waktu Penelitian	35
3.2 Jenis Penelitian	35
3.3 Instrumen Penelitian	35
3.4 Prosedur Penelitian	35
3.5 Studi Pustaka	37
3.6 Pengumpulan Data.....	38
3.7 Perancangan dan Analisis	41
3.7.1 Perancangan Praproses Data.....	41
3.7.2 Perancangan Algoritma VAE-CNN.....	44
3.7.3 Perancangan Alur Pengujian.....	47
3.7.3.1 <i>Flowchart</i>	47
3.7.3.2 <i>Data Flow Diagram (DFD)</i>	49
3.7.4 Analisis Data.....	52
BAB IV HASIL DAN PEMBAHASAN	76
4.1 Persiapan Data (<i>Data Preparation</i>).....	76
4.1.1 Penyaringan Data (<i>Data Filtering</i>).....	77
4.1.2 Pembersihan Data (<i>Data Cleaning</i>).....	79
4.2 Praproses Data (<i>Data Preprocessing</i>)	81
4.2.1 Pelabelan Data (<i>Data Labeling</i>)	83
4.2.2 Tokenisasi (<i>Tokenization</i>).....	83
4.2.3 Pengkodean Level Karakter (<i>Character-Level Encoding</i>)	84
4.2.4 Penyesuaian Panjang (<i>Padding</i>)	84
4.2.5 Pembentukan <i>Tensor 4D</i>	85
4.2.6 Segmentasi Data (<i>Data Segmentation</i>)	85

4.3 Pengembangan Model VAE-CNN	86
4.3.1 Model VAE-CNN dengan <i>Latent Dimension</i> 4.....	88
4.3.2 Model VAE-CNN dengan <i>Latent Dimension</i> 8.....	91
4.3.3 Model VAE-CNN dengan <i>Latent Dimension</i> 16.....	94
4.3.4 Analisis dan Pemilihan Model VAE-CNN Terbaik	96
4.4 Implementasi Sistem Pengujian.....	98
4.4.1 Antarmuka Pengguna (<i>User Interface</i>).....	98
4.4.2 Pengujian dengan Sampel URL	100
4.5 Perbandingan Model VAE-CNN dengan Model Lain.....	101
BAB V PENUTUP	103
5.1 Kesimpulan	103
5.2 Saran	103
DAFTAR PUSTAKA	105
LAMPIRAN.....	111



DAFTAR TABEL

Tabel 2.1 Kajian Literatur	12
Tabel 2.2 <i>Confusion Matrix</i>	34
Tabel 3.1 Transformasi Data	44
Tabel 4.1 Spesifikasi Lingkungan Uji Coba Google Colab	76
Tabel 4.2 <i>Hyperparameter</i> Terpilih	88
Tabel 4.3 Metrik Evaluasi (<i>Latent Dimension 4</i>)	91
Tabel 4.4 Metrik Evaluasi (<i>Latent Dimension 8</i>)	93
Tabel 4.5 Metrik Evaluasi (<i>Latent Dimension 16</i>)	96
Tabel 4.6 Perbandingan Metrik Evaluasi Model	98
Tabel 4.7 Hasil Pengujian pada Sistem	100
Tabel 4.8 Perbandingan Model VAE-CNN dengan Model Lain	101



DAFTAR GAMBAR

Gambar 1.1 Jumlah Serangan <i>Phishing</i> pada Q3 2022 - Q3 2024	2
Gambar 1.2 Data <i>Domain Abuse</i> Sepanjang Tahun 2024	3
Gambar 2.1 Skema <i>Phishing</i>	20
Gambar 2.2 Struktur URL	22
Gambar 2.3 Jaringan <i>Deep Learning</i>	23
Gambar 2.4 Logo Python	25
Gambar 2.5 Arsitektur <i>Variational Autoencoder</i> (VAE)	26
Gambar 2.6 Arsitektur <i>Convolutional Neural Network</i> (CNN)	26
Gambar 2.7 Representasi Skematis VAE-CNN	28
Gambar 3.1 Metodologi Penelitian.....	36
Gambar 3.2 PhishDataset	39
Gambar 3.3 <i>Tweet</i> dengan URL <i>Phishing</i> di Media Sosial X.....	39
Gambar 3.4 <i>Tweet Harvest</i>	40
Gambar 3.5 Inisialisasi Proses <i>Data Crawling</i> dengan <i>Tweet Harvest</i>	41
Gambar 3.6 URLScan.io <i>Phishing URL Feed</i>	41
Gambar 3.7 Alur Praproses Data.....	42
Gambar 3.8 Alur Proses Model VAE-CNN	45
Gambar 3.9 <i>Flowchart</i> Sistem Pengujian.....	48
Gambar 3.10 DFD Level 0	49
Gambar 3.11 DFD Level 1	49
Gambar 3.12 <i>Wireframe</i> Halaman Deteksi.....	50
Gambar 3.13 <i>Wireframe</i> Halaman Deteksi (<i>Toggle Sidebar</i>)	51
Gambar 3.14 <i>Wireframe</i> Input URL Manual	51
Gambar 3.15 <i>Wireframe</i> Input dalam <i>Batch</i>	52
Gambar 4.1 Dataset Data_bal - 20000 (<i>Unfiltered</i>)	77
Gambar 4.2 Dataset Data_bal - 20000 (<i>Filtered</i>)	78
Gambar 4.3 URLScan.io <i>Phishing Feed (Unfiltered)</i>	78
Gambar 4.4 URLScan.io <i>Phishing Feed (Filtered)</i>	79
Gambar 4.5 URL <i>Phishing</i> Hasil <i>Crawling (Uncleaned)</i>	79

Gambar 4.6 URL <i>Phishing</i> Hasil <i>Crawling</i> (<i>Cleaned</i>)	81
Gambar 4.7 Kurva <i>Loss</i> Pelatihan (<i>Latent Dimension</i> 4).....	89
Gambar 4.8 <i>Distribution of Reconstruction Loss</i> (<i>Latent Dimension</i> 4).....	90
Gambar 4.9 <i>Confusion Matrix</i> (<i>Latent Dimension</i> 4).....	90
Gambar 4.10 Kurva <i>Loss</i> Pelatihan (<i>Latent Dimension</i> 8).....	91
Gambar 4.11 <i>Distribution of Reconstruction Loss</i> (<i>Latent Dimension</i> 8).....	92
Gambar 4.12 <i>Confusion Matrix</i> (<i>Latent Dimension</i> 8).....	93
Gambar 4.13 Kurva <i>Loss</i> Pelatihan (<i>Latent Dimension</i> 16).....	94
Gambar 4.14 <i>Distribution of Reconstruction Loss</i> (<i>Latent Dimension</i> 16).....	95
Gambar 4.15 <i>Confusion Matrix</i> (<i>Latent Dimension</i> 16).....	96
Gambar 4.16 Perbandingan Kurva <i>Loss</i> Pelatihan Model.....	97
Gambar 4.17 Halaman Deteksi.....	98
Gambar 4.18 Halaman Deteksi (<i>Toggle Sidebar</i>)	99
Gambar 4.19 Input URL Manual.....	99
Gambar 4.20 Input URL dalam <i>Batch</i>	99

