



SKRIPSI

PERBANDINGAN METODE CNN-SVM DENGAN CNN- DECISION TREE UNTUK DETEKSI PHISING BERBASIS URL

REZA AMINULLAH

NPM 20081010234

DOSEN PEMBIMBING

Fetty Tri Anggraeny, S.Kom. M. Kom

Fawwaz Ali Akbar, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PERBANDINGAN METODE CNN-SVM DENGAN CNN- DECISION TREE UNTUK DETEKSI PHISING BERBASIS URL

REZA AMINULLAH

NPM 20081010234

DOSEN PEMBIMBING

Fetty Tri Anggraeny, S.Kom. M. Kom

Fawwaz Ali Akbar, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

LEMBAR PENGESAHAN

PERBANDINGAN METODE CNN-SVM DENGAN CNN-DECISSION TREE UNTUK DETEKSI PHISING
BERBASIS URL

Oleh :

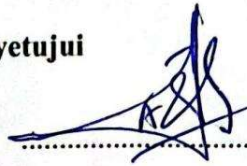
Reza Aminullah

NPM. 20081010234

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Padatanggal 10 Maret 2025

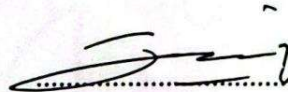
Menyetujui

Fetty Tri Anggraeny, S.Kom. M. Kom
NIP. 19820211 2021212 005



(Pembimbing I)

Fawwaz Ali Akbar, S.Kom, M.Kom
NIP. 19920317 2018031 002



(Pembimbing II)

Budi Nugroho, S.Kom, M.Kom
NIP. 198009072021211005



(Ketua Penguji)

Eva Yulia Puspaningrum, S.Kom., M.Kom
NIP. 19890705 2021212 002



(Anggota Penguji II)

Mengetahui

Dekan Fakultas Ilmu Komputer



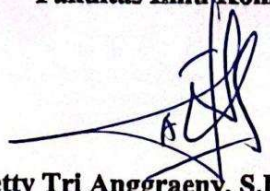
Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN
PERBANDINGAN METODE CNN- SVM DENGAN CNN - DECISION TREE UNTUK DETEKSI
PHISING BERBASIS URL

Oleh : Reza Aminullah
NPM. 20081010234

Menyetujui

Koorrdinator Prodi Informatika
Fakultas Ilmu Komputer



Fetty Tri Anggraeny, S.Kom. M. Kom
NIP. 19820211 2021212 005

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama : Reza Aminullah
NPM : 20081010234
Program : Sarjana(S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Tugas Akhir/Skripsi/Tesis/Disertasi* ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemuan indikasi plagiat pada Skripsi/Tesis/Desertasi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 23 Mei 2025

Yang Membuat pernyataan


Reza Aminullah
20081010234

ABSTRAK

Nama Mahasiswa / NPM : Rea Aminullah / 20081010234
Judul Skripsi : Perbandingan Metode CNN-SVM Dengan CNN-Decission Tree
Untuk Deteksi Phising Berbasis URL
Dosen Pembimbing : 1. Fetty Tri Anggraeny, S.Kom. M. Kom
2. Fawwaz Ali Akbar, S.Kom, M.Kom

URL phishing adalah bentuk kejahatan digital yang mencuri data pribadi pengguna internet dengan menyamar sebagai situs web tepercaya. Kata "phishing" berasal dari kata asing "fishing", yang merujuk pada cara penipuan yang digunakan untuk menarik target. Tujuan utama URL phishing adalah untuk memperoleh informasi penting seperti kata sandi, nomor kartu kredit, atau data pribadi lainnya dengan mengarahkan target ke situs web yang tampak sah. Untuk melindungi pengguna dari ancaman ini, diperlukan sistem deteksi dan pencegahan yang efektif untuk mengidentifikasi URL yang mencurigakan.

Studi ini mengusulkan untuk membandingkan dua model pembelajaran mesin hibrid, yaitu CNN-SVM dan CNN-Decision Tree, untuk mendeteksi serangan phishing berbasis URL. Dalam kedua model, jaringan saraf konvolusional (CNN) digunakan untuk melakukan ekstraksi fitur dari URL yang dianalisis. Fitur yang dihasilkan kemudian diklasifikasikan menggunakan mesin vektor pendukung (SVM) pada model pertama dan pohon keputusan pada model kedua. Hasil evaluasi menunjukkan bahwa pada skenario pertama, model CNN-SVM memiliki akurasi sebesar 0,835, tingkat recall sebesar 0,865, F1-Score sebesar 0,83, akurasi keseluruhan sebesar 82,86%, dan kehilangan akurasi sebesar 0,7020. Pada saat yang sama, model pohon keputusan CNN mencapai akurasi sebesar 0,86, tingkat ingatan sebesar 0,885, Skor F1 sebesar 0,865, akurasi keseluruhan sebesar 86,67%, dan kehilangan presisi sebesar 0,7436. Hasil penelitian menunjukkan bahwa kedua model memiliki kemampuan yang baik dalam mengidentifikasi URL phishing, dimana CNN-Decision Tree memiliki akurasi yang sedikit lebih baik pada seluruh skenario percobaan, dengan selisih sebesar 0,07%. Penelitian ini memberikan wawasan tentang efektivitas kedua pendekatan dalam mendeteksi ancaman phishing dan dapat menjadi dasar untuk pengembangan lebih lanjut di bidang keamanan siber.

Kata kunci: Deteksi *Phishing URL*, *CNN*, *SVM*, *Decission Tree*

ABSTRACT

Student Name / NPM : Reza Aminullah / 20081010234
Thesis Title : Comparison Of CNN-SVM and CNN-Decission Tree Methods
for URL Based Phising
Advisors : 1. Fetty Tri Anggraeny, S.Kom. M. Kom
2. Fawwaz Ali Akbar, S.Kom, M.Kom

Phishing URLs are a form of digital crime that masquerade as trusted websites and target Internet users to steal personal information. The term "phishing" is derived from the English word "fishing," which means "lure," and describes the deceptive tactics used to lure victims. The main purpose of phishing URLs is to obtain sensitive information such as passwords, credit card numbers, or other personal data by redirecting victims to seemingly legitimate websites. To protect users from this threat, an effective detection and prevention system that can detect suspicious URLs is required.

This study compares two hybrid machine learning models, namely CNN-SVM and CNN-Decision Tree, for detecting URL-based phishing attacks. In both models, convolutional neural networks (CNNs) are used to extract features from the analyzed URLs. The generated features are then classified using support vector machines (SVMs) in the first model and decision trees in the second model. The evaluation results show that the CNN-SVM model achieves a precision of 0.835, a recall of 0.865, and an F1-score of 0.83 in the first scenario, with an overall accuracy of 82.86% and a loss accuracy of 0.7020. The CNN-Decision Tree model achieved a precision of 0.86, a recall of 0.885, an F1 score of 0.865, an overall accuracy of 86.67%, and a loss accuracy of 0.7436. These results indicate that both models are good at detecting phishing URLs. CNN-Decision Tree showed slightly better accuracy in all experimental scenarios, leading by 0.07%. This study provides insight into the effectiveness of both methods in detecting phishing threats and can serve as a foundation for further developments in network security.

Keywords: URL Phishing Detection, CNN, SVM, Decision Tree.

KATA PENGANTAR

Dengan memanjatkan doa ku memohon rahmat dan karunia Tuhan YME Allah SWT yang telah memberikan kesehatan, kekuatan dan kemampuan pada penulis sehingga penulis dapat menyelesaikan dan menyelesaikan penulisan makalah dengan judul **“Perbandingan Pendekatan CNN-SVM dengan CNN-Decision Tree untuk Deteksi Phishing Berbasis URL”**.

Terima kasih juga penulis sampaikan kepada ibu semoga memberi kesehatan dan kelancaran pekerjaan IbuFetty Tri Anggraeny, S.Kom, M. Kom. sebagai instruktur utama yang telah memberikan waktu dan kesempatan kepada para penulis secara pribadi dalam menulis makalah ini serta memberikan arahan dan motivasi semangat untuk menyelesaikan makalah ini, serta bantuan dari segala aspek in moral, spiritual dan material, maka penulis ucapkan kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer UniversitasPembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Fetty Tri Anggraeny, S.Kom. M. Kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran “ Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom. M. Kom selaku pembimbing pertama saya yang sudah mengantarkan saya mulai jenjang sempro hingga jenjang semhas.
4. Bapak Fawwaz Ali Akbar, S.Kom, M.Kom selaku pembimbing kedua saya yang sudah mendidik saya untuk tidak mudah puas dengan hasil yang hanyaa dilaksanakan sekali perrcobaan.
5. Dosen-dosen Program Studi Informatika.
6. Kedua orang tua yang selalu mendukung walaupun banyak kekurangan dan juga telah membantu membayar ukt walau sudah melebihi masa studi.
7. Mas Odong selaku PM waktu bekerja di KPU Provinsi Jawa Timur yang selalu membantu meringankan jadwal kerja saya.
8. Pak Eko selaku rekan kerja di KPU Provinsi Jawa Timur yang sudah mengajari bagaimana cara membagi waktu untuk bekerja dan juga menyelesaikan studi.
9. Mas Patrick selaku rekan kerja di KPU Provinsi Jawa Timur yang selalu mendukung ketika ada tekanan waktu kerja dan juga tuntutan studi yang harus diselesaikan.
10. Mahendrra Wisnu Wardana selaku rekan kerrja d KPU Provinsi Jawa Timur

maupun teman seangkatan yang telah membimbing saya sejaksempro hingga semhas.

11. Iffan Maulana selaku rekan kerja di KPU Provinsi Jawa Timur maupun teman seangkatan Yang sudah mendukung saya.
12. Rekan Rekan Media Center KPU Provinsi Jawa Timur yang sudah memberi dukungan selama ini.
13. Reza Nur Hamza sebagai teman seangkatan yang sudah datang mendukung waktu ujian

Penulis mengakui berbagai kekurangan dalam penyusunan artikel ini. Oleh karena itu, penulis sangat mengharapkan kritik dan saran dari semua pihak guna meningkatkan kualitas penulisan artikel ini. Meskipun terdapat keterbatasan, penulis berharap laporan ini dapat memberikan manfaat bagi semua pihak, khususnya penulis sendiri.

Surabaya, _____

Reza Aminullah

DAFTAR ISI

LEMBAR PENGESAHAN.....	Error! Bookmark not defined.
LEMBAR PERSETUJUAN	Error! Bookmark not defined.
SURAT PERNYATAAN ORISINALITAS	Error! Bookmark not defined.
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Identifikasi Masalah	3
1.4 Tujuan Penelitian.....	3
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu	5
2.2 Phishing	6
2.3 Machine learning	7
2.4 Metode Support Vector Machine(SVM).....	8
2.5 Convolutional Neural Network (CNN)	10
2.6 Decission Tree	13
2.7 Metode Evaluasi.....	15
2.8 Analisis Hasil.....	17
BAB III DESAIN DAN IMPLEMENTASI SISTEM	19
3.1 Waktu dan Tempat Penelitian.....	19
3.2 Alat dan Bahan.....	19
3.3 Sumber Data.....	19
3.4 Pemodelan	20
3.5 Skenario pengujian	28
BAB IV PENGUJIAN DAN ANALISA	31
4.1 Deskripsi Data Set.....	31

4.2	Model <i>CNN</i>	36
4.3	Klasifikasi <i>SVM</i>	38
4.4	Klasifikasi Decision Tree	40
4.5	Evaluasi matriks pengujian	41
4.6	Analisis Evaluasi Peforma Pengujian	49
BAB V KESIMPULAN DAN SARAN		59
5.1	Kesimpulan	59
5.2	Saran	60
DAFTAR PUSTAKA		63
LAMPIRAN		67

DAFTAR GAMBAR

Gambar 2. 1 Batas Keputusan dataset	9
Gambar 2. 2Arsitektur CNN [14]	11
Gambar 2. 3Convolution Layer	12
Gambar 2. 4 Pooling Layer	12
Gambar 2. 5Konsep Pohon Keputusan.....	13
Gambar 3. 1 Flowchart Praproses.....	20
Gambar 3. 2 Flowchart proses CNN	22
Gambar 3. 3 Flowchart proses CNN ke SVM.....	24
Gambar 3. 4 Flowchart proses CNN ke Decission Tree.....	26
Gambar 4. 1 data set phishingURL	31
Gambar 4. 2 kode pembersihan data dan lematisasi.....	32
Gambar 4. 3 kode tokenisasi.....	34
Gambar 4. 4 kode dari padding dan reshaping	35
Gambar 4. 5 kode dari normalisasi.....	36
Gambar 4. 6 kode input layer dan embedding layer.....	36
Gambar 4. 7 kode convolutional layer	37
Gambar 4. 8 kode pooling layer	37
Gambar 4. 9 kode flattening layer	37
Gambar 4. 10 kode dense layer	37
Gambar 4. 11 kode ekstraksi fitur	38
Gambar 4. 12 kode Inisialisasi Model SVM	38
Gambar 4. 13 kode pelatihan model SVM	38
Gambar 4. 14 kode prediksi data ujii menggunakan SVM	38
Gambar 4. 15 kode pengambilan label asli data uji.....	39
Gambar 4. 16 kode evaluasi perfnoma SVM	39
Gambar 4. 17 kode inisialisasi model Decission Tree.....	40
Gambar 4. 18 kode pelatihan model Decission Tree.....	40
Gambar 4. 19 kode prediksi data uji Decission Tree.....	40
Gambar 4. 20 kode evaluasi kinerja model Decission Tree	41
Gambar 4. 21 Gravik Evaluasi Model CNN - SVM	53

DAFTAR TABEL

Tabel 3. 1 Skenario pengujian	28
Tabel 4. 1 Perbandingan data sebelum dan sesudah cleansing serta lemmatisasi	33
Tabel 4. 2 Perbandingan data sebelum dan sesudah tokenisasi	34
Tabel 4. 3 Perbandingan data sebelum dan sesudah padding	35
Tabel 4. 4 Perbandingan data sebelum dan sesudah reshaping	35
Tabel 4. 5 Evaluasi SVM scenario pertama	41
Tabel 4. 6 Evaluasi Decission Tree scenario pertama	42
Tabel 4. 7 Evaluasi SVM scenario kedua	42
Tabel 4. 8 Evaluasi Decission Tree scenario kedua	43
Tabel 4. 9 Evaluasi SVM scenario ketiga	43
Tabel 4. 10 Evaluasi Decission Tree scenario ketiga	43
Tabel 4. 11 Evaluasi SVM scenario keempat	44
Tabel 4. 12 Evaluasi Decission Tree scenario keempat	44
Tabel 4. 13 Evaluasi SVM scenario kelima	45
Tabel 4. 14 Evaluasi Decission Tree scenario pertama	45
Tabel 4. 15 Evaluasi SVM scenario keenam	46
Tabel 4. 16 Evaluasi Decission Tree scenario keenam	46
Tabel 4. 17 Evaluasi SVM scenario ketujuh	46
Tabel 4. 18 Evaluasi Decission Tree scenario ketujuh	47
Tabel 4. 19 Evaluasi SVM scenario kedelapan	47
Tabel 4. 20 Evaluasi Decission Tree scenario kedelapan	48
Tabel 4. 21 Evaluasi SVM scenario kesembilan	48
Tabel 4. 22 Evaluasi Decission Tree scenario kesembilan	48
Tabel 4. 23 Hasil evaluasi setiap skenario CNN-SVM	49
Tabel 4. 24 Hasil evaluasi setiap skenario Decission Tree	49
Tabel 4. 25 Tabel perbandingan split data	57
Tabel 4. 26 Tabel perbandingan Epoch Layer	57