



SKRIPSI

IMPLEMENTASI *FASTER* R-CNN DAN OCR DENGAN PENDEKATAN *NATURAL LANGUAGE PROCESSING* UNTUK DETEKSI IKLAN JUDI *ONLINE*

ANDRY SYVA MALDINI
NPM 21083010085

DOSEN PEMBIMBING

Wahyu Syaifullah J. S, S.Kom., M.Kom.
Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng.

KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025



SKRIPSI

**IMPLEMENTASI *FASTER* R-CNN DAN OCR
DENGAN PENDEKATAN *NATURAL LANGUAGE
PROCESSING* UNTUK DETEKSI IKLAN JUDI
*ONLINE***

ANDRY SYVA MALDINI
NPM 21083010085

DOSEN PEMBIMBING

Wahyu Syaifullah J. S, S.Kom., M.Kom.
Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**



SKRIPSI

**IMPLEMENTASI *FASTER* R-CNN DAN OCR
DENGAN PENDEKATAN *NATURAL LANGUAGE
PROCESSING* UNTUK DETEKSI IKLAN JUDI
*ONLINE***

ANDRY SYVA MALDINI
NPM 21083010085

DOSEN PEMBIMBING

Wahyu Syaifullah J. S, S.Kom., M.Kom.
Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng.

**KEMENTERIAN PENDIDIKAN, TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2025**

LEMBAR PENGESAHAN

IMPLEMENTASI *FASTER R-CNN* DAN OCR DENGAN PENDEKATAN *NATURAL LANGUAGE PROCESSING* UNTUK DETEKSI IKLAN JUDI ONLINE

Oleh:
ANDRY SYVA MALDINI
NPM. 21083010085

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 16 Juli 2025:

Wahyu Syaifullah J. S, S.Kom., M.Kom.
NIP. 198608252021211003



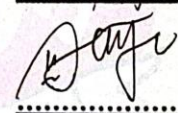
(Pembimbing I)

Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT.,
IPU., Asean. Eng.
NIP. 198012052005011002



(Pembimbing II)

Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 199507232024061002



(Ketua Penguji)

Alfan Rizaldy Pratama, S.Tr.T., M.Tr.Kom.
NIP. 199906062024061001



(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 196811261994032001

LEMBAR PERSETUJUAN

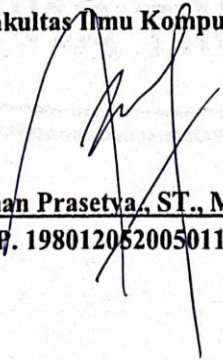
**IMPLEMENTASI *FASTER R-CNN* DAN OCR DENGAN PENDEKATAN
NATURAL LANGUAGE PROCESSING UNTUK DETEKSI IKLAN JUDI
*ONLINE***

Oleh:
ANDRY SYVA MALDINI
NPM. 21083010085

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**


Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng.
NIP. 198012052005011002

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : Andry Syva Maldini
NPM : 21083010085
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi 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 di kemudian hari ditemukan indikasi plagiat pada Skripsi 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, 16 Juli 2025
Yang Membuat Pernyataan,



ANDRY SYVA MALDINI
NPM. 21083010085

ABSTRAK

Nama Mahasiswa / NPM : Andry Syva Maldini / 21083010085
Judul Skripsi : Implementasi *Faster R-CNN* Dan OCR Dengan Pendekatan *Natural Language Processing* Untuk Deteksi Iklan Judi *Online*
Dosen Pembimbing : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom.
2. Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng

Penyebaran iklan judi online melalui media sosial semakin marak dan kerap dilakukan secara terselubung menggunakan *watermark* visual dan eufemisme teks, sehingga menyulitkan proses moderasi konten secara manual. Penelitian ini bertujuan mengembangkan sistem deteksi otomatis untuk mengidentifikasi konten iklan judi online dengan pendekatan *end-to-end* berbasis kecerdasan buatan. Metodologi yang digunakan mencakup pelatihan model deteksi objek menggunakan *Faster R-CNN* (ResNet-50 + FPN), ekstraksi teks dengan *Optical Character Recognition* berbasis TrOCR, dan klasifikasi teks menggunakan model BERT. Dataset pelatihan diperoleh melalui proses anotasi manual dan augmentasi dengan *Roboflow*. Evaluasi model menunjukkan bahwa *Faster R-CNN* mencapai mAP@0.50 sebesar 98,1%, model TrOCR menghasilkan *Character Error Rate* (CER) 4,6% dan *Word Error Rate* (WER) 29%, sedangkan model BERT untuk klasifikasi teks memperoleh akurasi sebesar 99% dengan *precision* dan *recall* tinggi pada kedua kelas. Ketiga komponen ini diintegrasikan dalam *pipeline* berbasis *Flask*, memungkinkan pengguna mengunggah gambar atau video pendek untuk diproses dan dianalisis secara otomatis. Hasil penelitian menunjukkan bahwa sistem mampu mendeteksi dan mengklasifikasikan iklan judi online, termasuk yang tersamarkan, secara efektif dan efisien, serta dapat digunakan sebagai alat bantu moderasi konten di *platform* digital.

Kata kunci : Deteksi Objek, *Faster R-CNN*, TrOCR, *BERT*, Iklan Judi *Online*, Kecerdasan Buatan, Moderasi Konten

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Andry Syva Maldini / 21083010085
Thesis Title : *Implementation of Faster R-CNN and OCR with a Natural Language Processing Approach for Detecting Online Gambling Advertisements*
Advisor : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom.
2. Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng

The proliferation of online gambling advertisements on social media has increasingly adopted covert strategies such as visual watermarks and textual euphemisms, making manual content moderation ineffective. This research aims to develop an automated detection system for identifying online gambling advertisements using an end-to-end artificial intelligence approach. The methodology involves object detection using Faster R-CNN (ResNet-50 + FPN), text extraction using Transformer-based Optical Character Recognition (TrOCR), and semantic classification using BERT-based Natural Language Processing. The dataset was collected and annotated manually, followed by augmentation using Roboflow. Evaluation results show that the Faster R-CNN model achieved a mAP@0.50 of 98.1%, TrOCR recorded a Character Error Rate (CER) of 4.6% and a Word Error Rate (WER) of 29%, while the BERT classification model achieved 99% accuracy, with high precision and recall for both classes. These components were integrated into a Flask-based web pipeline, allowing users to upload images or short videos for automatic detection and classification. The system effectively detects both explicit and obfuscated gambling advertisements and provides a practical tool to support content moderation efforts on digital platforms.

Keywords: *Object Detection, Faster R-CNN, TrOCR, BERT, Online Gambling Ads, Artificial Intelligence, Content Moderation*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah serta karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Implementasi *Faster R-CNN* Dan OCR Dengan Pendekatan *Natural Language Processing* Untuk Deteksi Iklan Judi Online”** sebagai syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Prodi Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Dari penyusunan Tugas Akhir ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU, selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya., ST., MT., IPU., Asean. Eng selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional “Veteran” Jawa Timur serta Dosen Pembimbing Dua.
4. Bapak Wahyu Syaifullah J. S, S.Kom., M.Kom., selaku Dosen Pembimbing Satu, yang telah memberikan bimbingan dan bantuan dalam proses penyusunan Skripsi ini.
5. Bapak dan Ibu dosen Program Studi Sains Data UPN “Veteran” Jawa Timur yang telah memberikan ilmu pengetahuan serta kontribusi berharga dalam penyelesaian skripsi ini.
6. Kedua orang tua beserta seluruh anggota keluarga yang senantiasa memberikan doa, dukungan, serta bantuan baik secara emosional maupun materiel kepada penulis.
7. Rekan-rekan Sains Data Angkatan 2021 yang selalu memberikan dorongan semangat dan dukungan selama penulis menyelesaikan penyusunan Tugas Akhir ini.
8. Terakhir, penulis ingin mengucapkan terima kasih kepada satu sosok yang selama ini diam-diam berjuang tanpa henti, seorang laki-laki sederhana

dengan impian yang tinggi, namun sering kali sulit ditebak isi pikiran dan hati. Terima kasih kepada penulis karya ini yaitu diriku sendiri, Andry Syva Maldini. Seorang pemuda di usianya yang kini menginjak 23 tahun, yang dikenal keras kepala namun terkadang sifatnya masih seperti anak kecil pada umumnya. Terima kasih telah turut hadir di dunia ini, telah bertahan sejauh ini, dan terus berjalan melewati segala tantangan yang semesta hadirkan. Terima kasih karena tetap berani menjadi dirimu sendiri. Aku bangga atas setiap langkah kecil yang kau ambil, atas semua pencapaian yang mungkin tak selalu dirayakan orang lain. Walau terkadang harapanmu tidak sesuai dengan apa yang semesta berikan, tetaplah belajar menerima dan mensyukuri apapun yang kamu dapatkan. Jangan pernah lelah untuk tetap berusaha, berbahagialah di manapun kamu berada. Rayakan apapun dalam dirimu dan jadikan dirimu bersinar di manapun tempatmu bertumpu.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 16 Juli 2025

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
DAFTAR LAMPIRAN	xix
DAFTAR NOTASI.....	xxi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu	9
2.2. Dasar Teori.....	13
2.2.1. <i>Artificial Intellegence</i>	13
2.2.2. <i>Machine Learning</i>	15
2.2.3. <i>Deep Learning</i>	16
2.2.4. <i>Convolutional Neural Network</i>	17
2.2.5. <i>Arsitektur Faster R-CNN</i>	19
2.2.6. <i>Arsitektur ResNet-50</i>	22
2.2.7. <i>Optical Character Recognition</i>	23
2.2.8. <i>Natural Language Processing</i>	25
2.2.9. <i>Bidirectional Encoder Representations from Transformers</i>	27

2.2.10.	<i>Confusion Matrix</i>	30
2.2.11.	<i>Mean Average Precision (mAP)</i>	33
2.2.12.	<i>Box Loss</i>	34
2.2.13.	<i>Class Loss</i>	35
2.2.14.	<i>Python</i>	36
2.2.15.	<i>Google Colaboratory</i>	36
2.2.16.	<i>Roboflow</i>	36
2.2.17.	<i>Flask</i>	37
BAB III DESAIN DAN IMPLEMENTASI SISTEM		39
3.1.	Variabel Penelitian dan Sumber Data	39
3.2.	Langkah Analisis.....	40
3.2.1.	Pengumpulan dan Anotasi Data	42
3.2.2.	<i>Preprocessing</i> Data	44
3.2.3.	Pelatihan Model <i>Faster R-CNN</i>	45
3.2.4.	Ekstraksi Gambar dari Deteksi (<i>Cropping</i>).....	47
3.2.5.	Implementasi OCR (TrOCR)	47
3.2.6.	Implementasi NLP (BERT).....	48
3.2.7.	Pembuatan <i>end to end Pipeline</i>	50
3.2.8.	Implementasi GUI <i>Flask</i>	51
3.2.9.	Spesifikasi Perangkat	52
3.3.	Langkah Analisis.....	53
BAB IV HASIL DAN PEMBAHASAN.....		57
4.1.	Informasi Dataset	57
4.2.	Pra-Pemrosesan Data	59
4.3.	Proses Pemodelan.....	72
4.4.	Evaluasi Model.....	79
4.5.	Implementasi Sistem Terintegrasi.....	84
4.6.	Hasil Pengujian	99
4.7.	Hasil <i>User Interface</i>	101
BAB V PENUTUP		107
5.1.	Kesimpulan	107
5.2.	Saran.....	108

DAFTAR PUSTAKA	109
LAMPIRAN.....	115

Halaman ini sengaja dikosongkan

DAFTAR GAMBAR

Gambar 2.1 Diagram Venn AI, ML, dan DL	14
Gambar 2.2 Arsitektur CNN	18
Gambar 2.3 Arsitektur <i>Faster</i> R-CNN.....	20
Gambar 2.4 Model Asitekur ResNet-50.....	23
Gambar 2.5 Arsitektur NLG & NLU [29]	25
Gambar 2.6 Arsitektur <i>Transformers</i>	27
Gambar 2.7 Proses <i>Pre-training</i> dan <i>Fine-tuning</i> BERT.....	28
Gambar 2.8 Visualisasi <i>bounding box</i> dalam rumus IoU	34
Gambar 3.1 Diagram Alir Penelitian	41
Gambar 3.2 Diagram Alir GUI <i>Flask</i>	51
Gambar 3.3 <i>Wireframe</i> Halaman <i>Home</i>	53
Gambar 3.4 <i>Wireframe</i> Halaman <i>Detect</i>	54
Gambar 3.5 <i>Wireframe</i> Halaman <i>About</i>	55
Gambar 4.1 Perbedaan sebelum (kiri) dan sesudah dianotasi (kanan).....	60
Gambar 4.2 Contoh Hasil Augmentasi Rotasi pada Citra Dataset.....	61
Gambar 4.3 Konfigurasi Augmentasi dan Split Dataset di <i>Roboflow</i>	61
Gambar 4.4 Contoh hasil pemotongan.....	64
Gambar 4.5 Visualisasi <i>WordCloud</i> untuk Kategori Judi	69
Gambar 4.6 Visualisasi <i>WordCloud</i> untuk Kategori Bukan Judi	70
Gambar 4.7 <i>Output</i> pelatihan <i>Faster</i> R-CNN	80
Gambar 4.8 Luaran Evaluasi Model NLP	83
Gambar 4.9 Bagian Utama Halaman <i>Home</i>	101
Gambar 4.10 Bagian Teknologi dan Footer pada Halaman <i>Home</i>	102
Gambar 4.11 Antarmuka Unggah Media pada Halaman <i>Detect</i>	102
Gambar 4.12 Halaman Hasil Analisis	103
Gambar 4.13 Bagian Awal Halaman <i>About</i>	103
Gambar 4.14 Penjelasan Detail Teknologi pada Halaman <i>About</i>	104
Gambar 4.15 Visualisasi Cara Kerja <i>DetectAI</i>	104
Gambar 4.16 Profil Pengembang pada Halaman <i>About</i>	105

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	9
Tabel 2.2 <i>Confusion Matrix</i>	30
Tabel 3.1 Ringkasan Dataset Primer	39
Tabel 3.2 Dataset Sekunder untuk <i>Transfer Learning</i>	40
Tabel 3.3 Jumlah Gambar Dataset Sekunder	43
Tabel 3.4 Parameter Pelatihan Model	45
Tabel 3.5 Parameter Pelatihan Model TrOCR	47
Tabel 3.6 Parameter Pelatihan Model BERT	49
Tabel 3.7 Spesifikasi Perangkat Penelitian	52
Tabel 4.1 Ringkasan Format dan Jenis Data pada Tiap Tahap Model.....	58
Tabel 4.2 Contoh struktur file CSV untuk TrOCR	65
Tabel 4.3 Tabel N-Gram Kategori Judi.....	69
Tabel 4.4 Tabel N-Gram Kategori Bukan Judi	70
Tabel 4.5 Parameter Pelatihan Model TrOCR	77
Tabel 4.6 Parameter Pelatihan Model NLP.....	78
Tabel 4.7 Hasil Evaluasi Model Faster R-CNN.....	80
Tabel 4.8 Evaluasi Model TrOCR per <i>Epoch</i>	82
Tabel 4.9 Hasil Evaluasi Model NLP	84
Tabel 4.10 Hasil Pengujian Sistem pada 20 Sampel Video	99
Tabel 4.11 Hasil Pengujian Sistem pada 20 Sampel Gambar	100

Halaman ini sengaja dikosongkan

DAFTAR LAMPIRAN

Lampiran 1 <i>Letter of Acceptance</i>	115
Lampiran 2 Kode Program.....	116

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

- H_x : Fungsi pemetaan (*mapping*) yang dipelajari oleh blok jaringan konvensional.
- R_x : Fungsi residual ($H(x)-x$) yang dipelajari oleh blok residual pada arsitektur ResNet.
- S : Jumlah *grid cell* dalam gambar
- B : Jumlah prediktor *bounding box* per *grid cell*
- λ_{coord} : Parameter bobot yang mengatur signifikansi dari *box loss* dalam total *loss*.
- 1_{ij}^{obj} : Fungsi indikator yang bernilai 1 jika objek terdeteksi pada *box loss* dalam total *loss*.
- (x_i, y_i) : Koordinat titik tengah (*center point*) dari *bounding box* prediksi.
- $(\hat{x}_i, \hat{y}_i)$: Koordinat titik tengah (*center point*) dari *bounding box ground truth*.
- (w_i, h_i) : Ukuran lebar (*width*) dan tinggi (*height*) dari *bounding box* prediksi.
- $(\hat{w}_i, \hat{h}_i)$: Ukuran lebar (*width*) dan tinggi (*height*) dari *bounding box ground truth*.
- $P_i(c)$: Probabilitas prediksi untuk kelas c pada grid ke- i .
- $\hat{P}_i(c)$: Label sebenarnya (*ground truth*) untuk kelas c pada grid ke- i (bernilai 1 atau 0).

Halaman ini sengaja dikosongkan