

**EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN
MENGUNAKAN TESSERACT**



**NATHAN GILBERT PASARIBU
2110511039**

**S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
JAKARTA
2025**

**EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN
MENGUNAKAN TESSERACT**

**NATHAN GILBERT PASARIBU
2110511039**

Skripsi

Sebagai salah satu syarat untuk melaksanakan
penelitian oleh mahasiswa pada

Program Studi Informatika

**S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA
JAKARTA
2025**

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Nathan Gilbert Pasaribu

NIM : 2110511039

Tanggal : 27 Mei 2025

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 27 Mei 2025

Yang Menyatakan



Nathan Gilbert Pasaribu

LEMBAR PERSETUJUAN

Yang bertanda tangan di bawah ini:

Nama : Nathan Gilbert Pasaribu
NIM : 2110511039
Program Studi : S-1 Informatika
Judul Skripsi : EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI
VETERAN MENGGUNAKAN TESSERACTION

Dinyatakan telah memenuhi syarat dan menyetujui untuk mengikuti ujian sidang skripsi/tugas akhir.

Jakarta, April 2025

Menyetujui,

Dosen Pembimbing I,



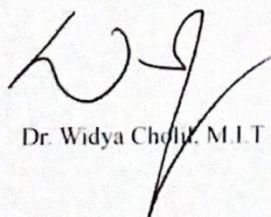
Indra Permana Solihin, S.Kom., M.Kom

Dosen Pembimbing II,



Novi Trisman Hadi, S.Pd., M.Kom

Mengetahui,
Koordinator Program Studi,



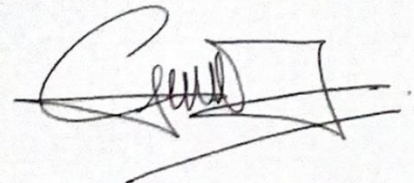
Dr. Widya Cholita, M.I.T

PERNYATAAN MENGENAI PROPOSAL SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa proposal skripsi dengan judul “EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN MENGGUNAKAN TESSERACT” adalah karya saya dengan arahan dari dosen 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 proposal ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Universitas Pembangunan Nasional Veteran Jakarta.

Jakarta, 27 Mei 2025

A handwritten signature in black ink, appearing to read 'Nathan Gilbert Pasaribu', with a large, stylized flourish extending to the right.

Nathan Gilbert Pasaribu
2110511039

LEMBAR PENGESAHAN

Judul : EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN
MENGGUNAKAN TESSERACT
Nama : Nathan Gilbert Pasaribu
NIM : 2110511039
Program Studi : S1 Informatika

Disetujui oleh :

Penguji 1:

Dr. Didit Widiyanto, S.Kom, M.Si.

Penguji 2:

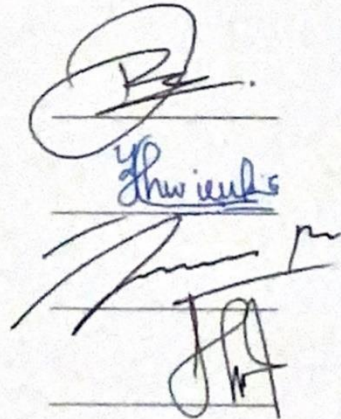
Kharisma Wiati Gusti, M.T.

Pembimbing 1:

Indra Permana Solihin, S.Kom, M.Kom.

Pembimbing 2:

Novi Trisman Hadi, S.Pd., M.Kom.



Diketahui oleh:

Koordinator Program Studi:

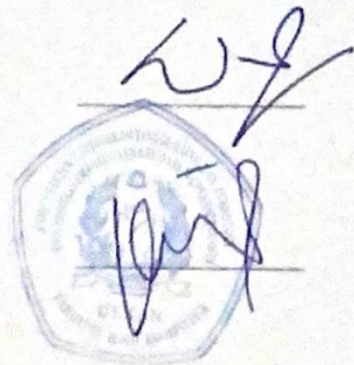
Dr. Widya Cholil, M.IT

NIP. 221112080

Dekan Fakultas Ilmu Komputer:

Prof. Dr. Ir. Supriyanto, M.Sc., IPM.

NIP. 197605082003121002



Tanggal Ujian Tugas Akhir:

27 Mei 2025

SKRIPSI : EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN MENGUNAKAN TESSERACT

Nathan Gilbert Pasaribu

ABSTRAK

Penelitian ini bertujuan untuk mengukur performa sistem Optical Character Recognition (OCR) dalam mengekstraksi data dari Kartu Tanda Penduduk (KTP). Sistem yang dikembangkan dievaluasi berdasarkan tiga aspek utama, yaitu tingkat akurasi hasil ekstraksi, waktu pemrosesan, dan efisiensi dibandingkan metode input manual. Hasil pengujian menunjukkan bahwa sistem OCR mampu mencapai tingkat akurasi rata-rata sebesar 90%. Waktu rata-rata yang dibutuhkan untuk memproses satu gambar KTP adalah 1,68 detik, dengan waktu tercepat 1,26 detik dan waktu terlama 2,93 detik. Selain itu, penggunaan OCR dapat mereduksi waktu proses input data hingga 97% dibandingkan dengan metode manual. Namun, akurasi sistem bervariasi tergantung pada elemen data yang diekstrak, dengan elemen RT/RW memiliki tingkat keberhasilan terendah sebesar 70%. Berdasarkan hasil tersebut, disarankan untuk mengeksplorasi model OCR alternatif seperti EasyOCR atau TrOCR, serta meningkatkan dataset dan metode praproses untuk memperbaiki akurasi ekstraksi.

Kata Kunci: OCR, KTP, ekstraksi data, Tesseract, akurasi, efisiensi waktu

*THESIS : EFFICIENCY OF PERSONAL DATA ENTRY IN VETERAN
INFORMATION SYSTEM USING TESSERACT*

Nathan Gilbert Pasaribu

ABSTRACT

This study aims to evaluate the performance of an Optical Character Recognition (OCR) system in extracting data from Indonesian identity cards (KTP). The developed system was assessed based on three main aspects: extraction accuracy, processing time, and efficiency compared to manual data entry. The results show that the OCR system achieved an average accuracy rate of 90%. The average processing time for a single KTP image was 1.68 seconds, with the fastest time recorded at 1.26 seconds and the slowest at 2.93 seconds. Additionally, the implementation of OCR reduced data input time by up to 97% compared to manual methods. However, extraction accuracy varied across different data fields, with RT/RW achieving the lowest success rate at 70%. Based on these findings, it is recommended to explore alternative OCR models such as EasyOCR or TrOCR, as well as improve datasets and preprocessing methods to enhance extraction accuracy.

Keywords: OCR, KTP, data extraction, Tesseract, accuracy, time efficiency

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas kasih dan penyertaan-Nya sehingga skripsi berjudul “EFISIENSI ENTRI DATA DIRI SISTEM INFORMASI VETERAN MENGGUNAKAN TESSERACT” dapat diselesaikan. Skripsi ini disusun dalam rangka memperoleh gelar Sarjana Komputer (S.Kom) pada program studi S1 Informatika dan diharapkan dapat memberikan dampak bagi lingkungan akademik Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.

Penulis mengucapkan terima kasih kepada:

1. Keluarga atas dukungan dan doa dalam menyelesaikan skripsi.
2. Prof. Dr. Ir. Supriyanto, M.Sc., IPM selaku dekan Fakultas Ilmu Komputer UPN Veteran Jakarta.
3. Dr. Widya Cholil, M.I.T selaku Koordinator Program Studi S1 Informatika
4. Bapak Indra Permana Solihin, S.Kom., M.Kom selaku dosen pembimbing pertama.
5. Bapak Novi Trisman Hadi, S.Pd., M.Kom selaku dosen pembimbing kedua.
6. Teman-teman sejawat S-1 Informatika yang mendukung selalu.

Penulis menyadari skripsi yang telah disusun ini jauh dari sempurna. Oleh karena itu, sangat diharapkan kritik dan saran yang membangun bagi skripsi ini. Semoga karya ini dapat bermanfaat bagi perkembangan ilmu pengetahuan.

Jakarta, 27 Mei 2025

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS.....	iii
LEMBAR PERSETUJUAN	iv
PERNYATAAN PELIMPAHAN HAK CIPTA	v
LEMBAR PENGESAHAN	vi
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat Penelitian.....	5
1.4.1 Tujuan penelitian.....	5
1.4.2 Manfaat penelitian.....	5
1.5 Sistematika Penulisan.....	5
BAB II. TINJAUAN PUSTAKA	7
2.1 Penelitian Terdahulu	7
2.2 Landasan Teori	8
2.2.1 YOLO.....	8
2.2.2 Tesseract-OCR	11
2.2.3 <i>Ground truth</i>	18
BAB III. METODE PENELITIAN.....	21
3.1 Identifikasi Masalah.....	21
3.2 Studi Literatur	21
3.3 Pengumpulan Data	21
3.4 <i>Training</i> dan Evaluasi <i>Bounding box</i>	22
3.5 Pembuatan <i>Ground truth</i>	22
3.6 <i>Finetune</i> dan Evaluasi Model OCR	22
3.7 Pembuatan tampilan <i>dummy</i> Sisfovet	22
3.8 Membangun <i>backend</i>	23

3.9 Implementasi OCR.....	23
3.10 Jadwal Penelitian	23
BAB IV. HASIL DAN PEMBAHASAN.....	24
4.1 <i>Training</i> dan Evaluasi <i>Bounding box</i>	24
4.2 Pembuatan <i>Ground truth</i>	28
4.3 <i>Finetune</i> dan evaluasi model OCR	29
4.4 Pembuatan tampilan <i>dummy</i> Sisfovet	31
4.5 Membangun <i>backend</i>	34
4.6 Implementasi fitur OCR.....	36
BAB V. PENUTUP	46
5.1 Kesimpulan.....	46
5.2 Saran	46
DAFTAR PUSTAKA.....	48
LAMPIRAN.....	49

DAFTAR GAMBAR

Gambar 1.1 Rekapitulasi Cavet	2
Gambar 2.1 Gambaran besar cara kerja YOLO (Redmon et al., 2015)	9
Gambar 2.2 CNN dalam YOLOv1 (Redmon et al., 2015).....	9
Gambar 3.1 Tahapan penelitian.....	21
Gambar 3.2 Contoh <i>bounding box</i> untuk KTP	22
Gambar 4.1 Tampilan CVAT	24
Gambar 4.2 Format file dalam folder untuk <i>training</i> YOLO.....	25
Gambar 4.3 Isi file data.yaml.....	25
Gambar 4.4 Kode <i>training</i> YOLO	26
Gambar 4.5 Metrik <i>Confusion</i>	26
Gambar 4.6 Metrik <i>loss</i> , <i>recall</i> , <i>precision</i> , dan <i>mAP</i>	27
Gambar 4.7 <i>Ground truth</i> berupa file teks dan gambar	28
Gambar 4.8 Kode python untuk pembuatan <i>ground truth</i>	29
Gambar 4.9 Kode untuk <i>training</i> Tesseract	29
Gambar 4.10 Hasil BCER dan BWER proses pelatihan.....	31
Gambar 4.11 Tampilan Sisfovet asli	32
Gambar 4.12 Kode <i>sidebar</i> tampilan dummy	32
Gambar 4.13 Kode <i>navbar</i> tampilan dummy.....	33
Gambar 4.14 Kode container tampilan dummy	33
Gambar 4.15 Tampilan dummy Sisfovet	33
Gambar 4.16 Kode python untuk praproses dengan metode HSV	34
Gambar 4.17 Hasil praproses gambar metode HSV	34
Gambar 4.18 Kode deteksi atribut dengan YOLO.....	35
Gambar 4.19 Kode pemotongan atribut KTP	35
Gambar 4.20 Kode OCR.....	36
Gambar 4.21 Kode pembersihan dan normalisasi teks hasil OCR	36
Gambar 4.22 Kode <i>testing</i> OCR	38
Gambar 4.23 Tingkat akurasi OCR per atribut	45

DAFTAR TABEL

Tabel 1.1 Waktu input manual.....	2
Tabel 2.1 Ringkasan penelitian terdahulu	7
Tabel 3.1 Jadwal penelitian.....	23
Tabel 4.1 Hasil proses OCR KTP	38
Tabel 4.2. Perbandingan waktu manual dan OCR	43