

**Optimisasi Pencocokan Kata dalam *Optical Character Recognition*
Menggunakan Metode *Hamming Distance***

SKRIPSI



Disusun Oleh:

Rakhammadhan Rizky Brillian

200602039

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH GRESIK**

2024

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT, karena berkat rahmat dan hidayahnya, penulis dapat menyelesaikan skripsi dengan judul “Optimisasi Pencocokan Kata dalam *Optical Character Recognition* Menggunakan Metode *Hamming Distance*” dengan lancar. Tujuan dari Skripsi ini digunakan untuk mempelajari cara pembuatan skripsi pada Universitas Muhammadiyah Gresik dan sebagai persyaratan menjadi Sarjana Komputer pada Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Gresik.

Selesainya skripsi ini tidak terlepas dari dukungan atau dorongan serta bimbingan, arahan, bantuan, saran dan doa dari semua pihak, oleh Karena itu pada kesempatan ini penulis akan menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, yang senantiasa memberikan kesehatan, kesabaran dan ketabahan penulis.
2. Ibu Nadhirotul Laily, S.Psi., M.Psi. P.hD., selaku Rektor Universitas Muhammadiyah Gresik.
3. Bapak Hanurnur Rosyid, ST., M.Kom. selaku Dekan Fakultas Teknik.
4. Ibu Henny Dwi Bhakti, S.Si., M.Si. selaku Kepala Program Studi Teknik Informatika. Serta selaku Dosen Wali.
5. Ibu Dr. Soffiana Agustin, S.Kom., M.Kom. selaku Pembimbing yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
6. Seluruh dosen informatika Universitas Muhammadiyah Gresik yang telah memberikan ilmu melalui mata kuliah yang diajarkan.
7. Bapak Kuncoro selaku ayah, ibu Wahyu Handari selaku ibu, dan seluruh saudara saya yang menjadi semangat penulis dalam mengerjakan penelitian.
8. Rekan-rekan seperjuangan Teknik Informatika Angkatan 2020 yang tidak dapat penulis sebutkan satu persatu.
9. Dan tidak lupa teman-teman yang telah memberikan semangat kepada penyusun dalam menyusun skripsi.

Penulis menyadari bahwa laporan skripsi ini masih banyak kekurangan dalam penulisan, serta masih jauh dari kesempurnaan. Oleh karena itu penulis mengharapkan dengan kelapangan hati penyusun menerima saran dan kritik yang sifatnya membangun bagi pembaca dan masyarakat umum serta masukan untuk menyempurnakan isi dari Skripsi ini. Dengan kerendahan hati penyusun mengucapkan banyak terima kasih dan semoga laporan ini bermanfaat bagi pembaca.

Gresik, 25 Maret 2024

Penyusun,

RAKHMADHAN RIZKY BRILLIAN

NIM. 200602039



ABSTRAK

Citra adalah representasi objek dua dimensi dari dunia visual, menyangkut berbagai macam disiplin ilmu yang mencakup seni, human vision, astronomi, teknik, dan sebagainya. Pada Penelitian ini, dilakukan penerapan Hamming Distance pada Optical Character Recognition menggunakan Matlab dan Python. Kata yang digunakan dalam peneliitian ini merupakan hasil tulis tangan dalam bahasa Indonesia, dimana Penulis dibagi dalam tiga kelompok dengan usia 10-15 tahun (kelompok A), 20-25 tahun (kelompok B) dan usia lebih dari 50 tahun (kelompok C). Penelitian dimulai dengan analisis citra dengan melakukan pra-pengolahan dan Ekstraksi kata menggunakan Optical Character Recognition (OCR). Hasil OCR ini kemudian dimasukkan ke pengenalan Hamming untuk dilakukan pengenalan Kata. Hasil penelitian menunjukkan kinerja OCR dalam mengenali kata memberikan akurasi tertinggi pada kelompok B yaitu 40%, sedangkan pengenalan kata menggunakan Hamming distance memberikan akurasi 85% pada kelas yang sama. Baik OCR maupun Hamming belum mampu mengenali kata dari tulisan tangan dengan baik pada kelompok A dan C.

Kata Kunci: optimasi, pencocokan kata, hamming distance, optical character recognition

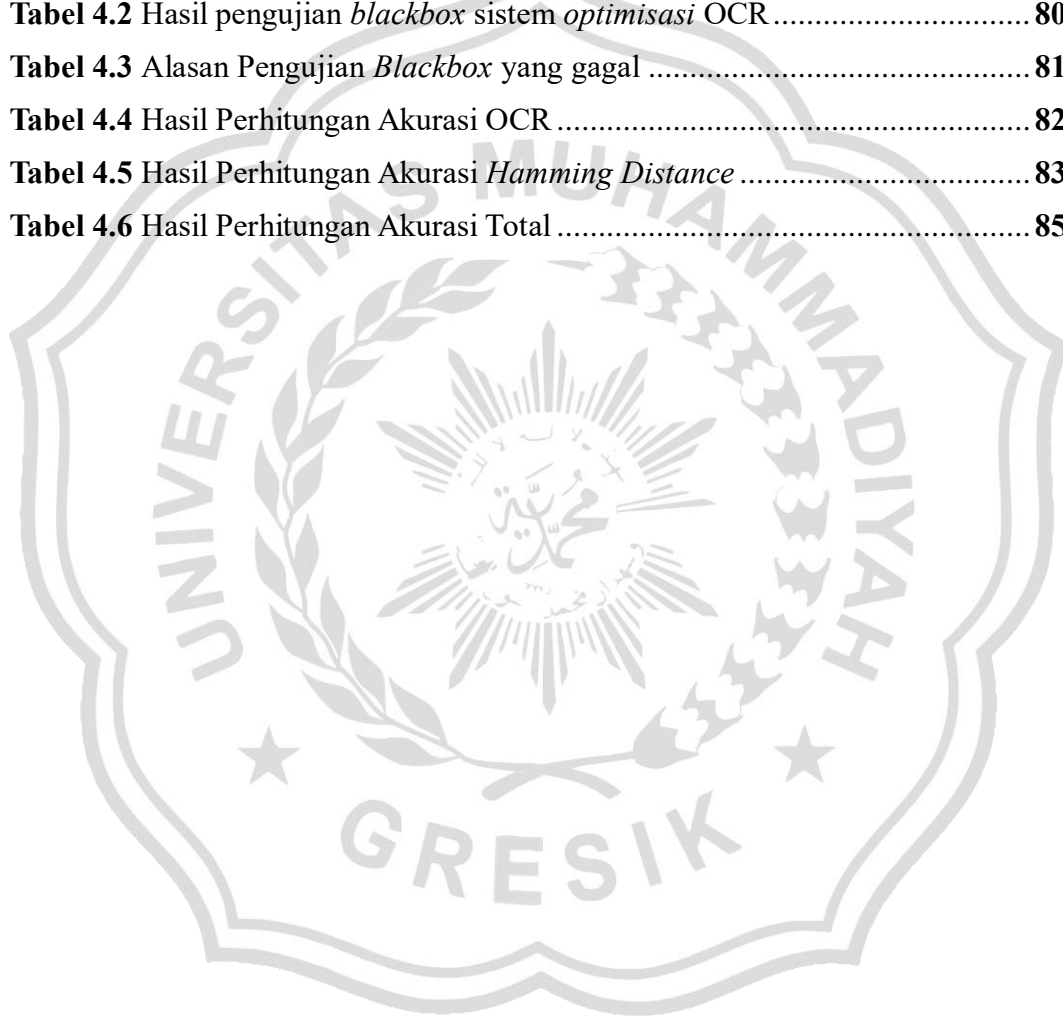
DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN	ii
PERNYATAAN KEASLIAN SKRIPSI.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR GAMBAR.....	viii
DAFTAR TABEL.....	x
BAB I.....	1
1.1 LATAR BELAKANG.....	1
1.2 PERUMUSAN MASALAH	3
1.3 TUJUAN PENELITIAN.....	3
1.4 MANFAAT PENELITIAN.....	3
1.5 BATASAN PENELITIAN	3
1.6 METODOLOGI PENELITIAN	4
1.6.1 STUDI LITERATUR.....	4
1.6.2 PENGUMPULAN DATA.....	4
1.6.3 ANALISIS DAN PERANCANGAN SISTEM.....	4
1.6.4 PENGUJIAN SISTEM	4
1.6.5 PEMBUATAN LAPORAN.....	4
1.7 SISTEMATIKA PENULISAN	5
BAB II	6
2.1 TINJAUAN PUSTAKA.....	6
2.2 LANDASAN TEORI.....	9
2.2.1 TULISAN.....	9
2.2.2 CITRA DIGITAL.....	11
2.2.3 PEMROSESAN CITRA DIGITAL	12
2.2.4 <i>CROPPING</i>	12
2.2.5 RGB	13
2.2.6 <i>GRAYSCALE</i>	14
2.2.7 <i>ENHANCEMENT</i>	15

2.2.8	<i>OPTICAL CHARACTER RECOGNITION</i>	15
2.2.9	MATLAB	16
2.2.10	<i>TESSERACT ENGINE</i>	23
2.2.11	<i>PYTHON</i>	23
2.2.12	PERSAMAAN KATA	24
2.2.13	<i>HAMMING DISTANCE</i>	27
BAB III	29
3.1	Analisis sistem.....	29
3.2	Perancangan Sistem	30
3.2.1	<i>Preprocessing</i> atau Pemrosesan awal	32
3.2.2	Ekstraksi Fitur	35
3.2.3	Persamaan Kata	55
3.3	Kebutuhan Perangkat Keras dan Perangkat Lunak	58
3.4	Perancangan Antarmuka Sistem	58
3.5	Skenario Pengujian	61
BAB IV	63
4.1	Implementasi	63
4.1.1.	Tahap <i>Contrast Stretching</i>	63
4.1.2.	Tahap <i>Grayscale</i>	64
4.1.3.	Tahap <i>Complemented</i>	65
4.1.4.	Tahap Ekstraksi Fitur OCR	66
4.1.5.	Tahap Persamaan Kata Hamming Distance	66
4.1.6.	Tampilan GUI.....	76
4.2	Pengujian sistem.....	79
4.2.1	Pengujian <i>Blackbox</i>	79
4.2.2	Pengujian Akurasi.....	82
BAB V	87
5.1.	Kesimpulan.....	87
5.2.	Saran	87
DAFTAR PUSTAKA	88

DAFTAR TABEL

Tabel 3.1 Hasil Ekstraksi Fitur (OCR)	54
Tabel 3.2 Hasil Proses 1, 2 dan 3 Persamaan Kata	56
Tabel 3.3 Hasil Pencocokan <i>Ground Truth</i>	57
Tabel 3.4 Rancangan Menu Sistem Pencocokan Kata Tulisan Tangan	59
Tabel 4.1 Informasi Citra Tulisan Tangan	63
Tabel 4.2 Hasil pengujian <i>blackbox</i> sistem <i>optimisasi</i> OCR	80
Tabel 4.3 Alasan Pengujian <i>Blackbox</i> yang gagal	81
Tabel 4.4 Hasil Perhitungan Akurasi OCR	82
Tabel 4.5 Hasil Perhitungan Akurasi <i>Hamming Distance</i>	83
Tabel 4.6 Hasil Perhitungan Akurasi Total	85



DAFTAR GAMBAR

Gambar 2.1 Warna <i>RGB</i>	14
Gambar 2.2 Jenis-jenis metode dalam persamaan kata	25
Gambar 2.3 Perbandingan Huruf	28
Gambar 3.1 Flowchart perancangan sistem	31
Gambar 3.2 Flowchart Tahapan Preprocessing	32
Gambar 3.3 Sebelum contrast (kiri) dan sesudah contrast stretching (kanan).....	33
Gambar 3.4 Citra RGB dan Kanal Penyusunnya	33
Gambar 3.5 Citra Grayscale dengan nilai piksel.....	34
Gambar 3.6 Hasil akhir tahapan preprocessing	35
Gambar 3.7 <i>Flowchart tesseract engine</i>	35
Gambar 3.8 <i>Flowchart Tahapan Page Layout Analysis</i>	37
Gambar 3.9 <i>Input gambar</i>	38
Gambar 3.10 Hasil Deteksi garis (kiri) dan gambar (kanan)	39
Gambar 3.11 Hasil Deteksi teks (kiri) dan titik henti teks (Tengah dan kanan) ..	40
Gambar 3.12 Deteksi Batas Kalimat (kiri) dan validasinya (Tengah) serta Deteksi kandidat batas antar kolom (kanan)	41
Gambar 3.13 Deteksi kolom (kiri), kolom kalimat (Tengah) dan kolom paragraf (kanan)	42
Gambar 3.14 <i>Flowchart recognize word pass</i>	44
Gambar 3.15 Sistem LSTM	51
Gambar 3.16 Sistem LSTM dengan <i>softmax</i>	53
Gambar 3.17 sistem LSTM <i>Tesseract</i>	53
Gambar 3.18 <i>Flowchart Tahapan Persamaan kata</i>	55
Gambar 3.19 Rancangan GUI <i>matlab</i>	58
Gambar 3.20 Rancangan GUI <i>Google colab</i>	60
Gambar 4.1 Citra awal (Kiri) dan Hasil <i>Contrast Stretching</i> (Kanan)	64
Gambar 4.2 citra sebelum <i>grayscale</i>	64
Gambar 4.3 citra hasil <i>grayscale</i>	65
Gambar 4.4 Hasil <i>grayscale</i> (Kiri) dan <i>complemented</i> (Kanan)	65
Gambar 4.5 Hasil <i>complemented</i> (Kiri) dan OCR(Kanan)	66
Gambar 4.6 Hasil teks OCR	66

