

**PENGENALAN KARAKTER HURUF KOREA (HANGEUL)  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)**

**TUGAS AKHIR**



**DINDA NADILA**

**1162001014**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS BAKRIE  
JAKARTA  
2023**

**PENGENALAN KARAKTER HURUF KOREA (HANGEUL)  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)**

**TUGAS AKHIR**

**Diajukan sebagai salah satu syarat untuk memperoleh gelar  
Sarjana Komputer**



**DINDA NADILA**

**1162001014**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK DAN ILMU KOMPUTER  
UNIVERSITAS BAKRIE  
JAKARTA**

**202**

## HALAMAN PERNYATAAN ORISINALITAS

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

Nama : Dinda Nadila

NIM : 1162001014

Tanda Tangan :



Tanggal : 28 Agustus 2023

## HALAMAN PENGESAHAN

Laporan Proposal Tugas Akhir ini diajukan oleh:

Nama : Dinda Nadila  
NIM : 1162001014  
Program Studi : Teknik Informatika  
Fakultas : Teknik dan Ilmu Komputer  
Judul Tugas Akhir : Pengenalan Karakter Huruf Korea (Hangeul)  
menggunakan Algoritma *Support Vector Machine* (SVM)

**Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie**

## DEWAN PENGUJI

Pembimbing 1 : Prof. Dr. Hoga Saragih, S.T., M.T.



Pembimbing 2 : Ihsan Ibrahim, S.T., M.T.



Penguji 1 : Guson Prasamuarso Kuntarto, S.T., M.Sc.

  
Agustus 30th, 2023

Penguji 2 : Berkah Iman Santoso, ST, MTI

  
August 29th, 2023 11:24 AM

Ditetapkan di : Jakarta

Tanggal : 28 Agustus 2023

## UCAPAN TERIMA KASIH

Segala puji atas kehadiran Allah SWT karena berkat segala Rahmat dan Karunia-Nya penulis dapat menyelesaikan Tugas Akhir yang berjudul Pengenalan Karakter Huruf Korea (Hangeul) menggunakan Algoritma *Support Vector Machine* (SVM). Begitu banyak pihak yang memberikan dukungan, turut membantu, memberikan masukan, nasihat, dan serta doa dalam penyusunan Tugas Akhir ini. Oleh karena itu, dengan segala hormat dan keserendahan hati, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Ihsan Ibrahim, S.T., M.T dan Bapak Prof. Dr. Hoga Saragih, S.T., M.T yang telah membantu dan telah meluangkan waktu serta mengarahkan pikiran dan tenaganya dalam proses pengerjaan dan penyusunan Tugas Akhir ini.
2. Bapak Guson Prasamuarso Kuntarto, S.T., M.Sc dan Berkah Iman Santoso, ST, MTI atas saran serta masukannya untuk kebaikan penulisan Laporan Tugas Akhir ini.
3. Papa, Mama, Abang, Kaka, dan Adik yang tidak pernah lelah untuk selalu memberikan doa, motivasi, dukungan, doa, nasihat, serta masukan.
4. Seluruh Dosen Program Studi Informatika yang telah memberikan ilmu dan pembelajaran serta pengalaman yang sangat bermanfaat selama masa perkuliahan di Universitas Bakrie.
5. Aiy dan Nida, terima kasih untuk semua dukungan, dan doa, serta semua hal yang telah kita hadapi dan jalanin bersama selama masa perkuliahan.
6. Guntur Setya Pramono yang senantiasa selalu memberikan dukungan, semangat, dan mengingatkan penulis.
7. Teruntuk Caweters a.k.a Cah Convers (Nia, Cellen, Yosi, Pingkan, Ibam, Yusron, dan Rendi) yang selalu ada dari jaman SMA, yang selalu mengisi hari-hari penulis dengan semangat dan canda tawa selama proses penyusunan penelitian.
8. Tim 'Berjuank', Aiy, Nida, dan Alif yang terus saling mengingatkan, membantu dan koordinasi selama proses penyelesaian Tugas Akhir ini.
9. Tim Miniherow (Cici, Nindy, Yaya, Dival, dan Yuni) yang selalu menjadi tim hore serta penyemangat.

10. Penyemangat jarak jauh anak-anakku Kai, Baekhyun, Sehun, Chanyeol, Kyungsoo, Suho, Chen, Xiumin, Lay, Bright, dan Win yang lagu-lagunya selalu menemani penulis selama penelitian.
11. Teman-teman Informatika Angkatan 2016 yang telah memberikan dukungan, semangat, motivasi serta kebersamaan dalam suka maupun duka selama perkuliahan, yang memberikan pengalaman dan pembelajaran selama masa perkuliahan di Universitas Bakrie.
12. Rekan kerja Generali Indonesia yang telah banyak memberikan pengalaman dan pembelajaran.

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Dinda Nadila  
NIM : 1162001014  
Program Studi : Informatika  
Fakultas : Teknik dan Ilmu Komputer

Dengan pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie Hak Bebas Royalti Noneksekutif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

**Pengenalan Karakter Huruf Korea (Hangeul) menggunakan Algoritma Support Vector Machine (SVM).** Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksekutif ini Universitas Bakrie berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*dataset*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Dengan demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta  
Pada Tanggal : 28 Agustus 2023

Yang Menyatakan,



Dinda Nadila

**PENGENALAN KARAKTER HURUF KOREA (HANGEUL)  
MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)**

**Dinda Nadila**

---

**ABSTRAK**

Pada era saat ini, minat penduduk Indonesia terhadap budaya dan Bahasa Korea terus berkembang pesat. Fenomena Hallyu Wave yang merajalela di Indonesia telah memicu minat besar dalam memahami budaya dan Bahasa Korea. Banyak dari mereka yang ingin mengunjungi Korea Selatan sebagai tujuan wisata, sehingga hal ini menjadi salah satu pendorong minat mereka untuk mempelajari Bahasa Korea. Efek ini tercermin dalam peningkatan minat dalam berbagai layanan pembelajaran Bahasa Korea di Indonesia, seperti kursus Bahasa Korea dan pemilihan jurusan Sastra Korea di perguruan tinggi. Namun, Bahasa Korea memiliki sistem penulisan non-romawi yang unik, yang menggunakan aksara khusus yang disebut Hangeul. Penulisan Hangeul memiliki aturan-aturan khusus terkait tata letak dan struktur karakter antara huruf vokal dan huruf konsonan. Beragam ketentuan penulisan ini, bersama dengan pola yang bervariasi, seringkali menjadi penyebab kesalahan dalam pengenalan huruf Hangeul. Oleh karena itu, penulis mengusulkan penelitian pengenalan karakter huruf Hangeul menggunakan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi *machine learning* terbaik dengan 2 (dua) teknik ekstraksi fitur yaitu zoning dan GLCM (*gray-level co-occurrence matrix*). Dikarenakan terbatasnya ketersediaan dataset pada wadah terbuka dan penelitian sebelumnya, penulis memutuskan untuk membuat dataset secara manual untuk pelatihan dan pengujian. Pelatihan dilakukan dengan 4,411 data menggunakan jenis font Gulim berukuran 14, dan hasil penelitian menunjukkan bahwa tingkat akurasi tertinggi diperoleh pada pengujian dengan ekstraksi fitur GLCM terhadap font Gulim berukuran 24, mencapai 68%.

**Kata Kunci:** Hangeul, *Hallyu Wave*, *Machine Learning*, *Support Vector Machine*, *Zoning*, *Gray-level co-occurrence matrix*.



**CHARACTER RECOGNITION TO KOREAN CHARACTER (HANGEUL)  
USING THE SUPPORT VECTOR MACHINE (SVM) ALGORITHM**

**Dinda Nadila**

---

**ABSTRACT**

*In the present era, the interest of the Indonesian population in Korean culture and language is rapidly growing. The widespread phenomenon of the Hallyu Wave in Indonesia has sparked a significant interest in understanding Korean culture and language. Many aspire to visit South Korea as a tourist destination, which has become a driving force behind their motivation to learn the Korean language. This heightened interest is evident in the increased demand for various Korean language learning services in Indonesia, such as Korean language courses and the selection of Korean literature majors at universities. However, the Korean language employs a unique non-Roman writing system known as Hangeul. The writing of Hangeul adheres to specific rules regarding the layout and structure of characters, including vowels and consonants. The diverse rules of writing, combined with varying patterns, often lead to challenges in accurately recognizing Hangeul characters. To address these challenges, the author proposes a research project focused on the recognition of Hangeul characters using the Support Vector Machine (SVM) algorithm, known as one of the best machine learning classification methods. This research incorporates two feature extraction techniques: zoning and GLCM (gray-level co-occurrence matrix). Due to limited availability of datasets in open repositories and previous research, the author decided to manually create a dataset for training and testing purposes. Training involved 4,411 data points using the Gulim font size 14, and the research results indicate that the highest accuracy rate was achieved during testing with GLCM feature extraction applied to Gulim font size 24, reaching 68%.*

**Keyword:** *Hangeul Characters, Hallyu Wave, Machine Learning, Support Vector Machine, Zoning, Gray-level co-occurrence matrix.*

## DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS.....	i
HALAMAN PENGESAHAN.....	ii
UCAPAN TERIMA KASIH.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	v
ABSTRAK.....	vi
<i>ABSTRACT</i> .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR RUMUS.....	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR SINGKATAN.....	xv
DAFTAR ISTILAH.....	xvi
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Ruang Lingkup Penelitian.....	4
1.4 Tujuan dan Manfaat Penelitian.....	4
1.4.1 Tujuan Penelitian.....	4
1.4.2 Manfaat Penelitian.....	4
1.5 Sistematika Penulisan.....	5
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>6</b>
2.1 Penelitian Terkait.....	6
2.2 Pengenalan Karakter ( <i>Character Recognition</i> ).....	9
2.3 Karakteristik Huruf Hangeul.....	10
2.4 Tahapan Pengenalan Karakter.....	12
2.4.1 <i>Pre-processing</i> .....	13
2.4.2 <i>Segmentation</i> .....	13
2.4.3 <i>Feature Extraction</i> .....	14
2.5 <i>Classification</i> .....	14
2.6 <i>Confusion Matrix</i> .....	18

BAB III METODOLOGI PENELITIAN .....	20
3.1 Pembuatan Data.....	21
3.2 <i>Pre-processing</i> .....	22
3.2.1 <i>Binarization</i> .....	22
3.2.2 <i>Complement</i> .....	23
3.2.3 <i>Normalization</i> .....	24
3.3 Segmentasi.....	25
3.4 <i>Feature Extraction</i> .....	26
3.4.1 <i>Zoning</i> .....	26
3.4.2 <i>Gray Level Co-Occurrence Matrix (GLCM)</i> .....	27
3.5 <i>Support Vector Machine</i> .....	28
3.6 Evaluasi .....	31
3.7 <i>Research Tools</i> .....	31
3.7.1 <i>Dataset</i> .....	31
BAB IV IMPLEMENTASI DAN HASIL PENELITIAN.....	34
4.1 Pembuatan Data.....	34
4.2 <i>Pre-processing</i> .....	35
4.2.1 <i>Binarization</i> .....	36
4.2.2 <i>Complement</i> .....	37
4.2.3 <i>Normalization</i> .....	37
4.3 Segmentasi.....	38
4.4 <i>Feature Extraction</i> .....	39
4.4.1 <i>Zoning</i> .....	39
4.4.2 <i>GLCM</i> .....	40
4.5 <i>Algoritma Support Vector Machine</i> .....	42
4.5.1 <i>Pelatihan Data</i> .....	42
4.5.2 <i>Pengujian Data</i> .....	43
4.6 <i>Confusion Matrix</i> .....	44
BAB V KESIMPULAN DAN SARAN.....	48
5.1 Kesimpulan.....	48
5.2 Saran .....	49
DAFTAR PUSTAKA .....	50

## DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terkait .....	8
Tabel 2.2 <i>Confusion Matrix</i> [19].....	18
Tabel 3.1 Tabel contoh penerapan binerisasi .....	22
Tabel 3.2 Tabel contoh penerapan <i>complement</i> .....	23
Tabel 3.3 Tabel contoh penerapan normalisasi .....	24
Tabel 3.4 Tabel contoh segmentasi karakter .....	26
Tabel 3.5 Fitur GLCM [23].....	28
Tabel 4.1 Jumlah Dataset yang digunakan.....	44
Tabel 4.2 <i>Implementation Result</i> menggunakan fitur GLCM.....	45
Tabel 4.3 <i>Implementation Result</i> menggunakan fitur <i>Zoning</i> .....	46

## DAFTAR GAMBAR

Gambar 2.1 Metode Pengenalan Karakter [6].....	9
Gambar 2.2 Alfabet Huruf Korea: Hangeul [8] .....	11
Gambar 2.3 Tata letak huruf suku kata Huruf Korea: Hangeul [9] .....	11
Gambar 2.4 Alur Sistem Pengenalan Karakter [11].....	13
Gambar 2.5 Contoh dua kelas dengan bidang pemisah [16].....	15
Gambar 2.6 Contoh dua kelas dengan <i>hyperplane</i> terbaik [16].....	16
Gambar 3.1 Kerangka Penelitian .....	20
Gambar 3.2 <i>Flowchart</i> Sistem .....	21
Gambar 3.3 <i>Flowchart</i> <i>Binarization</i> .....	22
Gambar 3.4 <i>Flowchart</i> <i>Complement</i> .....	23
Gambar 3.5 <i>Flowchart</i> <i>Normalization</i> .....	24
Gambar 3.6 <i>Flowchart</i> Segmentasi.....	25
Gambar 3.7 <i>Flowchart</i> <i>Zoning</i> .....	26
Gambar 3.8 <i>Flowchart</i> Pelatihan Data di SVM.....	28
Gambar 3.9 Tipe-tipe kernel pada SVM [25] .....	29
Gambar 3.10 <i>Flowchart</i> Pengujian Data di SVM.....	30
Gambar 3.11 Kumpulan kombinasi huruf.....	32
Gambar 3.12 Data Latih.....	32
Gambar 3.13 Kata dalam Bahasa Korea .....	33
Gambar 3.14 Data Uji .....	33
Gambar 4.1 Data Latih.....	34
Gambar 4.2 Data Uji .....	35
Gambar 4.3 Data sebelum proses Binerisasi.....	36
Gambar 4.4 Data sesudah proses Binerisasi .....	36
Gambar 4.5 Data hasil proses <i>Complement</i> .....	37
Gambar 4.6 Hasil gambar <i>pre-processing</i> .....	38
Gambar 4.7 Data sebelum segmentasi .....	38
Gambar 4.8 Hasil <i>pre-processing</i> dan segmentasi.....	38
Gambar 4.9 Hasil segmentasi tidak sempurna .....	39
Gambar 4.10 Hasil Fitur <i>Zoning</i> .....	40
Gambar 4.11 Hasil Fitur GLCM .....	41
Gambar 4.12 Grafik perbandingan fitur (2 kelas).....	42

## DAFTAR RUMUS

Persamaan 2.1 <i>Support Vector Machine</i> model <i>Decision Boundary</i> .....	16
Persamaan 2.2 Sampel Positif.....	16
Persamaan 2.3 Sampel Negatif.....	16
Persamaan 2.4 Nilai Margin.....	17
Persamaan 2.5 Optimasi Konstrain .....	17
Persamaan 2.6 Konstrain .....	17
Persamaan 2.7 Metode <i>Lagrange Multiplier</i> .....	17
Persamaan 2.8 <i>Accuracy</i> .....	18
Persamaan 2.9 <i>Precision</i> .....	19
Persamaan 2.10 <i>Recall</i> .....	19
Persamaan 2.11 <i>F-measure</i> .....	19
Persamaan 3.1 Kontras GLCM .....	26
Persamaan 3.2 Korelasi GLCM .....	26
Persamaan 3.3 Homogenitas GLCM .....	26
Persamaan 3.4 Energi GLCM .....	26

**DAFTAR LAMPIRAN**

Lampiran 1 Dataset .....	53
Lampiran 2 Grafik perbandingan Fitur GLCM.....	80
Lampiran 3 Grafik perbandingan Fitur Zoning.....	83

**DAFTAR KODE**

<i>Code 4.1 Binerisasi</i> .....	36
<i>Code 4.1 Complement</i> .....	37
<i>Code 4.3 Training</i> .....	42



**DAFTAR SINGKATAN**

SVM	<i>Support Vector Machine</i>
OCR	<i>Optical Character Recognition</i>
HCR	<i>Handwritten Character Recognition</i>
K-NN	<i>K-Nearest Neighbor</i>
GLCM	<i>Gray Level Co-occurrence Matrix</i>
RBF	<i>Radial Basic Function</i>
ICZ	<i>Image Centroid Zone</i>
ZCZ	<i>Zone Centroid and Zone</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
TN	<i>True Negative</i>
FN	<i>False Negative</i>
MLP-BP	<i>Multilayer Perceptron Neural Network Backpropagation Algorithm</i>
MLP-LM	<i>Multilayer Perceptron Neural Network, Algoritma Levenberg- Marquardt</i>
CNN	<i>Convolutional Neural Network</i>

## DAFTAR ISTILAH

<i>Accuracy</i>	Jumlah data terdeteksi dengan benar
<i>Binarization</i>	Proses mengubah gambar asli menjadi gambar biner dengan tujuan agar setiap piksel pada gambar memiliki nilai biner yaitu 1 dan 0
<i>Complement</i>	Pertukaran nilai 1 dan 0 pada gambar biner yang bertujuan agar objek utama pada gambar yang akan di proses memiliki nilai 1
<i>Confusion Matrix</i>	Metode yang sering digunakan untuk menghitung tingkat akurasi suatu Sistem Pendukung Keputusan atau Data Mining
<i>F-measure</i>	Digunakan untuk mendapatkan nilai hasil kombinasi dari perhitungan <i>recall</i> dan <i>precision</i>
<i>Lagrange Multiplier</i>	Sebuah metode untuk mencari nilai maksimum atau minimum suatu fungsi pada sistem
<i>Normalization</i>	Bertujuan untuk mengubah nilai intensitas piksel suatu citra tanpa mengubah informasi utama pada citra tersebut
<i>Pattern Matching</i>	Metode klasifikasi dengan cara mencocokkan pola
<i>Precision</i>	Tingkat rasio hasil prediksi benar positif dengan keseluruhan hasil yang deteksi positif
<i>Pre-processing</i>	Bertujuan untuk melakukan proses yang terfokus pada operasi menghilangkan noise gambar yang diinput atau pindai
<i>Recall</i>	Menunjukkan keberhasilan model dalam menemukan informasi dalam sistem
<i>Supervised Learning</i>	Salah satu teknik pembelajaran <i>machine learning</i> yang dalam implementasinya mesin melatih data yang sudah diberi label terlebih dahulu