

**Pengenalan Citra Aksara Jawa  
dengan Ekstraksi Fitur *Principal Component Analysis*  
dan Pengklasifikasi *Support Vector Machine***



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Komputer  
Pada Jurusan Ilmu Komputer/Informatika**

**Disusun oleh:**

**ANGGIT GUSTI NUGRAHENI**

**24010314120015**

**DEPARTEMEN ILMU KOMPUTER/INFORMATIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO**

**2019**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Anggit Gusti Nugraheni

NIM : 24010314120015

Judul : Pengenalan Citra Aksara Jawa dengan Ekstraksi Fitur *Principal Component Analysis* dan Pengklasifikasi *Support Vector Machine*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 18 Februari 2019



Anggit Gusti Nugraheni  
24010314120015

## HALAMAN PENGESAHAN

Judul : Pengenalan Citra Aksara Jawa dengan Ekstraksi Fitur *Principal Component Analysis*  
dan Pengklasifikasi *Support Vector Machine*

Nama : Anggit Gusti Nugraheni

NIM : 24010314120015

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **8 Februari 2019**.

Semarang, 18 Februari 2019

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika  
FSM UNDIP



Dr. Retno Kusumaningrum, S.Si., M.Kom.  
NIP. 198104202005012001

Panitia Penguji Skripsi  
Ketua,

Drs. Suhartono, M.Kom.  
NIP. 195504071983031003

## HALAMAN PENGESAHAN

Judul : Pengenalan Citra Aksara Jawa dengan Ekstraksi Fitur *Principal Component Analysis*  
dan Pengklasifikasi *Support Vector Machine*

Nama : Anggit Gusti Nugraheni

NIM : 24010314120015

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal **8 Februari 2019**.

Semarang, 18 Februari 2019

Dosen Pembimbing



Helmie Arif Wibawa, S.Si, M.Cs  
NIP. 197805162003121001

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas limpahan karunianya, sehingga kegiatan skripsi dan penelitian ini dapat berjalan dengan baik. Kegiatan ini dilaksanakan dalam rangka pemenuhan salah satu syarat kelulusan dan untuk mendapatkan gelar sarjana pada Departemen Ilmu Komputer / Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro.

Selama proses pengerjaan skripsi ini, penulis mendapatkan banyak bantuan, bimbingan serta dorongan dari berbagai pihak. Baik secara langsung maupun tidak langsung. Penulis mengucapkan terima kasih kepada semua pihak yang terlibat dalam pengerjaan skripsi ini, terutama:

1. Ibu Dr. Retno Kusumaningrum, S.Si., M.Kom, selaku ketua Departemen Ilmu Komputer / Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro
2. Bapak Helmie Arif Wibawa, S.Si, M.Cs, selaku koordinator skripsi Departemen Ilmu Komputer / Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro dan dosen pembimbing skripsi
3. Dosen penguji, bapak Drs. Suhartono, M.Kom, ibu Sukmawati Nur Endah, S.Si, M.Kom, dan bapak Helmie Arif Wibawa, S.Si, M.Cs
4. Keluarga dan teman-teman yang selalu mendukung selama proses penulisan skripsi ini.

Penulis menyadari masih terdapat kekurangan dalam penyajian dan penulisan laporan skripsi ini. Untuk itu, penulis mengharapkan kritik dan saran yang membangun agar dapat memberikan hasil yang lebih baik. Hasil penelitian ini, diharapkan dapat bermanfaat untuk penelitian selanjutnya terkait pengenalan aksara jawa dan kehidupan masyarakat.

Semarang, 18 Februari 2019

Penulis

## HALAMAN PERNYATAAN PESETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Anggit Gusti Nugraheni  
NIM : 24010314120015  
Program Studi : Informatika  
Departemen : Ilmu Komputer/Informatika  
Fakultas : Sains dan Matematika  
Jenis karya : Skripsi

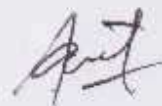
demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Noneksklusif (Non-exclusive RoyaltyFree Right)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

*Pengenalan Citra Aksara Jawa dengan Ekstraksi Fitur Principal Component Analysis dan Pengklasifikasi Support Vector Machine*

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/ formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 18 Februari 2019  
Yang menyatakan



Anggit Gusti Nugraheni  
24010314120015

## ABSTRAK

Bahasa Jawa merupakan salah satu bahasa lokal di Indonesia. Penulisan tradisional dalam Bahasa Jawa disebut aksara jawa. Sebagai huruf non-latin, aksara jawa memiliki cara penulisan dan bentuk yang unik. Penelitian ini menggunakan metode *principal component analysis* (PCA) dan *support vector machine* (SVM) untuk pengenalan citra aksara jawa. Metode PCA bertindak sebagai metode ekstraksi fitur, sementara SVM sebagai pengklasifikasi. Metode SVM yang digunakan pada penelitian ini adalah kernel *linier* dan strategi *one vs one*. Data yang digunakan sebanyak 27 kelas dengan masing-masing kelas berisi 3 citra uji dan 27 citra latih. Data tersebut dibagi menggunakan *k-fold cross validation* kedalam 10 *fold*. Penelitian ini melibatkan tiga skenario, yaitu PCA-SVM, *Thining*-PCA-SVM, dan *Zoning*-PCA-SVM. Hasil penelitian menunjukkan bahwa proses pengenalan aksara jawa dari kombinasi metode PCA dan SVM menghasilkan nilai yang berbeda untuk setiap skenario. Nilai rata-rata akurasi tertinggi dari metode PCA dan SVM diperoleh pada skenario *Zoning*-PCA-SVM, yaitu sebesar 52,3457% dengan *runtime* sebesar 187,528 detik.

**Kata kunci:** *principal component analysis*, *support vector machine*, aksara jawa, pengenalan citra

## ABSTRACT

Javanese is one of the local languages in Indonesia. Traditional writing in Javanese is called Javanese scripts. As a non-Latin letter, Javanese script has a unique font and writing method. The study was conducted to recognition Javanese script image by using principal component analysis (PCA) and support vector machine (SVM) methods. The PCA acted as a feature extraction method, while SVM acted as a classifier method. The SVM method used in the study was linear kernel and one vs one strategy. The data used were 27 classes, and each class consisted of 3 testing images and 27 training images. The data were divided by using k-fold cross validation into 10 fold. This study involved three tasks, that were PCA-SVM, Thining-PCA-SVM, and Zoning-PCA-SVM. Results showed that the Javanese script recognition process of the PCA and SVM method combination produced different values for the three skenarios. The highest accuracy average value of the PCA and SVM methods was gain from the Zoning-PCA-SVM skenario, that was 52.3457% with the runtime value 187.528 seconds.

**Key words:** principal component analysis, support vector machine, javanese scripts, image recognition

## DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PENGESAHAN .....	iv
KATA PENGANTAR .....	v
HALAMAN PERNYATAAN PESETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS .....	vi
ABSTRAK.....	vii
ABSTRACT .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN .....	xiii
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan dan Manfaat.....	3
1.4. Ruang Lingkup .....	3
1.5. Sistematika Penulisan.....	4
<b>BAB II LANDASAN TEORI.....</b>	<b>5</b>
2.1. Tinjauan Pustaka .....	5
2.2. Dasar Teori .....	7
2.2.1. <i>Principal componen analysis</i> .....	7
2.2.2. <i>Support vector machine</i> .....	11
2.2.3. <i>K-fold cross validation</i> .....	13
2.2.4. <i>Median filter</i> .....	13

<b>BAB III METODOLOGI PENELITIAN.....</b>	<b>16</b>
3.1. Pengumpulan Data.....	16
3.2. Prapengolahan .....	18
3.3. Ekstraksi Fitur .....	22
3.4. Pembagian Citra .....	23
3.5. Pelatihan dan Pengujian .....	24
3.6. Evaluasi .....	26
3.7. Analisis .....	27
3.8. Desain .....	29
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>35</b>
4.1. Proses Penelitian.....	35
4.2. Hasil Penelitian.....	37
<b>BAB V PENUTUP .....</b>	<b>47</b>
5.1. Kesimpulan.....	47
5.2. Saran .....	47
<b>DAFTAR PUSTAKA .....</b>	<b>48</b>
<b>LAMPIRAN-LAMPIRAN.....</b>	<b>50</b>

## DAFTAR GAMBAR

Gambar 2.1 Representasi Citra dari Matriks ke Vektor .....	8
Gambar 3.1 Garis Besar Penyelesaian Masalah .....	16
Gambar 3.2 Contoh Hasil <i>Scan</i> Data Citra.....	18
Gambar 3.3 Citra dengan <i>Noise</i> .....	19
Gambar 3.4 Penghilang <i>Noise</i> .....	19
Gambar 3.5 Pemotongan dan Penskalaan Citra.....	20
Gambar 3.6 Perubahan Fitur Warna Citra .....	20
Gambar 3.7 <i>Zoning</i> .....	21
Gambar 3.8 <i>Thining</i> .....	21
Gambar 3.9 Proses Ekstraksi Fitur PCA.....	22
Gambar 3.10 Pembagian Citra dengan <i>K-fold cross validation</i> .....	23
Gambar 3.11 Pelatihan dengan SVM .....	24
Gambar 3.12 Pengujian SVM (Strategi <i>Multiclass One Vs One</i> ).....	25
Gambar 3.13 Struktur Data.....	28
Gambar 3.14 DFD Level 0 .....	28
Gambar 3.15 DFD Level 1 .....	29
Gambar 3.16 Antarmuka Simulator.....	30
Gambar 3.17 Diagram Alir Proses Prapengolahan.....	31
Gambar 3.18 Diagram Alir Mengubah Citra Menjadi <i>Grayscale</i> .....	31
Gambar 3.19 Diagram Alir Proses Pembagian Citra.....	32
Gambar 3.20 Diagram Alir Mencari Indeks Latih .....	33
Gambar 3.21 Diagram Alir Proses Ekstraksi Fitur.....	33
Gambar 3.22 Diagram Alir Proses Pelatihan.....	34
Gambar 3.23 Diagram Alir Proses Pengujian .....	34
Gambar 4.1 Pengenalan Skenario PCA-SVM.....	36
Gambar 4.2 Pengenalan Skenario <i>Thining</i> -PCA-SVM.....	36
Gambar 4.3 Pengenalan Skenario <i>Zoning</i> -PCA-SVM .....	36
Gambar 4.4 Hasil Pengenalan .....	37
Gambar 4.5 Sampel Citra Setiap Kelas .....	42

## DAFTAR TABEL

Tabel 2.1 Perbandingan dengan Penelitian Lain .....	6
Tabel 3.1 Data Citra.....	17
Tabel 3.2 Pembagian Kelas .....	26
Tabel 3.3 <i>Software Requirement Specification</i> .....	27
Tabel 4.1 <i>Confusion Matrix</i> PCA-SVM.....	38
Tabel 4.2 <i>Confusion Matrix Thining</i> -PCA-SVM .....	39
Tabel 4.3 <i>Confusion Matrix Zoning</i> -PCA-SVM .....	40
Tabel 4.4 Contoh Hasil Citra Prapengolahan .....	43
Tabel 4.5 Akurasi dan <i>Runtime</i> .....	44
Tabel 4.6 <i>Confusion Matrix</i> Skenario PCA-SVM <i>Fold</i> ke-1 .....	46

## DAFTAR LAMPIRAN

Lampiran 1 Citra Hasil <i>Scan</i> .....	51
--	----

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Bahasa Jawa merupakan bahasa khas daerah Jawa Tengah. Aksara jawa merupakan bentuk kuno dalam penulisan Bahasa Jawa. Saat ini, aksara jawa masih dipelajari sebagai materi lokal pada sekolah-sekolah di Jawa Tengah. Aksara jawa terdiri dari 20 huruf dasar beserta pasangannya, 10 penulisan angka, 5 aksara swara, 8 aksara murda beserta pasangannya, 5 aksara rekan beserta pasangannya, dan beberapa sandangan (Piyono, 2005). Cara penulisan aksara jawa berbeda dengan tulisan latin. Penulisan dapat ditulis menggantung di batas atas garis tulis.

Setiap aksara memiliki pembacaan seperti satu suku kata, contohnya aksara Ha (ꦲꦲ), Na (ꦤꦲ), Ca (ꦕꦲ). Untuk memberi bunyi yang lain, diperlukan penambahan sandangan. Sebagai contoh pada aksara Ha (ꦲꦲ), jika ditambah sandangan i (ꦲꦲꦶ) menjadi Hi (ꦲꦲꦶ), ditambah sandangan e (ꦲꦲꦺ) menjadi He (ꦲꦲꦺ), ditambah sandangan è (ꦲꦲꦺꦴ) menjadi Hè (ꦲꦲꦺꦴ), atau ditambah sandangan o (ꦲꦲꦺꦴꦺ) menjadi Ho (ꦲꦲꦺꦴꦺ).

Aksara jawa memiliki banyak tata cara penulisan dan keunikannya. Dengan segala keunikan yang dimiliki aksara jawa ini, menjadikannya bahan yang menarik dalam penelitian pengenalan pola. Adanya penelitian-penelitian terkait aksara jawa, diharapkan pada masa mendatang dapat dimanfaatkan sebagai metode pembelajaran dalam dunia pendidikan dan melestarikan budaya.

Penelitian tentang pengenalan pola aksara jawa telah dilakukan dengan berbagai metode, diantaranya *K-Nearest neighbour* (Mukhooyar, 2015), *self organizing maps* (Hidayat & Shofa, 2016), *Back Propagation* (Nurmila, Sugiharto, & Sarwoko, 2010), dan *learning vektor quantization* (Mafrur, Andestoni, Ahdi, Fajri, & Muhantini, 2011). Tetapi, penelitian-penelitian tersebut baru sebatas pengenalan terhadap huruf-huruf dasar aksara jawa. Ketika dilakukan penambahan sandangan atau pasangan, maka akan memunculkan masalah baru seperti bertambahnya nilai fitur yang diperlukan dan kelas

yang dikenali. Untuk itu, diperlukan metode ekstraksi fitur dan pengklasifikasi yang lebih baik, seperti metode *principal componen analysis* dan *support vector machine*.

Metode *principal componen analysis* (PCA) merupakan metode yang cukup populer dalam bidang pengenalan. Terdapat beberapa penelitian yang menggunakan metode ini. Diantaranya ekstraksi fitur untuk deteksi kanker paru-paru dari gambar CT scan (Ada & Kaur, 2013); serta penelitian pengenalan citra wajah (Pratiwi, 2014), (Fu, 2015), dan (Budi, Suma'inna, & Maulana, 2016). Pada penelitian-penelitian tersebut, PCA mampu memberikan akurasi pengenalan citra di atas 70%. Percobaan yang dilakukan oleh Abdu Rakhman Syakhala, dkk (2015), menunjukkan PCA mampu memberikan waktu pengenalan yang lebih cepat dan hasil pengenalan yang lebih baik dibanding metode *hiden markov model* (HMM). Pada penelitian tersebut, PCA memiliki *runtime* antara 1-1,5 detik dengan akurasi tertinggi sebesar 86,6%. Sedangkan HMM, *runtime*-nya bisa mencapai 2-7,5 detik dan akurasi tertinggi 77,7%.

Selain PCA, terdapat juga metode *support vector machine* (SVM) yang dinilai baik dalam berbagai penelitian. SVM merupakan metode yang biasa digunakan dalam mengklasifikasi antara dua kelas. Seperti penelitian untuk diagnosa hepatitis (Munawarah, Soesanto, & Faisal, 2016); dan untuk prediksi ketepatan waktu kelulusan mahasiswa (Pratama, Wihandika, & Ratnawati, 2018). Namun, tak sedikit juga penelitian yang telah dilakukan untuk menggunakan SVM dalam permasalahan *multiclass*. Seperti penelitian mengklasifikasi 5 jenis penyakit Skizofrenia (Perdana, Furqon, & Indriati, 2018); dan mengklasifikasi *multiclass* sentimen (Liu, Bi, & Fan, 2017).

Beberapa penelitian yang telah dilakukan dengan SVM, terbukti cukup memuaskan. Percobaan yang pernah dilakukan untuk membandingkan SVM dengan metode lain, menunjukkan SVM sebagai pengklasifikasi yang lebih unggul. Seperti percobaan yang membandingkan antara metode SVM, Fuzzy-KNN, dan *binary decision tree* (BDT) pada klasifikasi detak jantung (Hasanah, M, Pratama, & Cholissodin, 2016), serta percobaan membandingkan SVM dengan *deep recurrent neural network* (RNN) sebagai penganalisis sentimen pada review hotel (Al-Smadi, Qawasmeh, Al-Ayyoub, Jararweh, & Gupta, 2017). SVM memiliki nilai akurasi yang lebih baik dibanding Fuzzy-KNN dan BDT, serta lebih unggul dari *deep* RNN dalam hal akurasi serta bobot rata-rata presisi dan sensitifitas (skor F1). Selain itu, SVM

dinilai sebagai pengklasifikasi yang baik meskipun data yang digunakan memiliki dimensi tinggi.

Pada penelitian ini, akan dilakukan percobaan untuk melihat kinerja kombinasi metode PCA dan SVM dalam mengenali citra aksara jawa. PCA akan bertindak sebagai ekstraksi fitur, sedangkan SVM sebagai pengklasifikasi. Kinerja kombinasi kedua metode ini akan dilihat dari perhitungan akurasi dan *runtime*.

## 1.2. Rumusan Masalah

Bagaimana kinerja dari kombinasi PCA sebagai ekstraksi fitur dan SVM sebagai pengklasifikasi, dalam melakukan pengenalan citra aksara jawa?

## 1.3. Tujuan dan Manfaat

Penelitian ini bertujuan untuk meneliti kemampuan kombinasi metode ekstraksi fitur PCA dan pengklasifikasi SVM dalam mengenali aksara jawa. Sedangkan manfaat yang dirasakan dari penelitian ini, berupa:

- a. Menambah wawasan terkait bidang pengenalan citra.
- b. Menjadi acuan dalam kegiatan penelitian mendatang, terkait aksara jawa.

## 1.4. Ruang Lingkup

Kegiatan penelitian skripsi ini dibatasi dengan ruang lingkup sebagai berikut:

1. Dari 20 huruf dasar aksara jawa, hanya digunakan 9 huruf (ha, na, ca, ka, da, ta, sa, wa, dan pa), dilihat dari segi kemiripannya.
2. Setiap huruf dasar yang dipilih, diberi 3 variasi yang berbeda. Menggunakan bentuk aslinya (tanpa sandangan), ditambah dengan sandangan i (wulu) dan penambahan sandangan e (pepet).
3. Proses pengenalan aksara jawa menggunakan metode *k-fold cross-validating* dengan *fold* sebanyak 10.
4. Strategi *multiclass* SVM yang digunakan berupa strategi *one vs one*.
5. Kemampuan PCA dan SVM dilihat melalui tingkat akurasi dan *runtime*.

## **1.5. Sistematika Penulisan**

Laporan skripsi ini dibuat dengan sistematika penulisan sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini dijabarkan latar belakang pemilihan masalah, rumusan masalah, tujuan dan manfaat penelitian, ruang lingkup permasalahan yang diambil, serta sistematika penulisan laporan.

### **BAB II LANDASAN TEORI**

Bab ini menjabarkan penelitian-penelitian serupa dimasa lampau dan teori-teori dasar yang digunakan dalam pelaksanaan penelitian.

### **BAB III METODOLOGI PENELITIAN**

Bab ini menjelaskan metode dan menganalisis rancangan yang digunakan dalam penelitian. Dimulai dari menjabarkan metode pengumpulan data, prapengolahan, ekstraksi fitur, pembagian data, pelatihan dan pengujian, analisis, serta desain.

### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini membahas proses dan hasil-hasil penelitian dari kegiatan skripsi, meliputi proses yang dilakukan pada setiap skenario serta hasil yang didapatkannya.

### **BAB V PENUTUP**

Bab ini berisi kesimpulan yang didapat dari hasil penelitian dan saran untuk perancangan yang dapat dilakukan pada penelitian mendatang.