

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Indonesia memiliki beragam suku dan budaya yang tersebar di seluruh daerah nusantara. Setiap suku memiliki ciri khas masing-masing yang unik. Aksara merupakan salah satu ciri khas dari keragaman suku di Indonesia. Menurut Pusat Bahasa Departemen Pendidikan Nasional dalam Kamus Bahasa Besar Indonesia (2008), Aksara merupakan sistem lambang grafis yang digunakan untuk berkomunikasi yang sedikit banyaknya mewakili ajaran. Pada zaman sekarang keberadaan aksara memudahkan masyarakat untuk berkomunikasi secara tidak langsung.

Aksara merupakan ucapan atau perkataan dalam bentuk tulisan (Kumalasari, 2010). Di Indonesia telah digunakan berbagai aksara seperti aksara Palawa yang dikenal dengan Aksara Jawa, Sunda, Arab, Bali dan yang terakhir dikenal dengan aksara Latin yang berasal dari Barat (Suastika, 2008). Menurut Kozok (2009) secara garis besar aksara nusantara dapat dikelompokkan menjadi beberapa kelompok besar diantaranya Aksara Hanacaraka yang berada di Jawa, Sunda dan Bali, Aksara Batak yang berada di Toba, Simalungun dan Karo, dan Aksara Ulu yang berada di Kerinci, Rejang, Lampung dan Serawi.

Aksara Sunda merupakan salah satu sejarah di Indonesia dalam hal aksara atau penulisan. Aksara sunda merupakan salah satu peninggalan sejarah asli masyarakat sunda, sejak abad ke 5 dan menjadi bahasa penulisan pada saat itu (Baidillah et al., 2008). Aksara Sunda merupakan hasil karya ortografi masyarakat Sunda. Secara umum lambang Aksara Sunda dapat dikelompokkan ke dalam enam kelompok yaitu aksara swara, aksara *Ngalagena*, aksara khusus, aksara rarangken, pasangan, dan angka (Baidillah et al., 2008).

Aksara Sunda terdiri atas 7 aksara vokal atau aksara swara dan 23 aksara konsonan atau aksara *ngalagena* (Farhan, Atmaja, & Aulia, 2017). Aksara *ngalagena* adalah tulisan yang secara silabis dianggap dapat melambangkan bunyi fonem konsonan dan dapat berperan sebagai sebuah kata maupun suku kata yang bisa menempati posisi awal, tengah, maupun akhir sebuah kata (Supriadi, 2009).

Aksara sunda merupakan salah satu kebudayaan yang telah berusia cukup lama yaitu selama lebih dari 16 abad, hal tersebut tampak pada prasasti-prasasti dari zaman itu yang sebagian besar telah dibicarakan oleh Kern (1917) dalam buku yang berjudul *Verspreide Geschriften; Inscripties van den Indischen Archipel* (Baidillah et al., 2008). Banyak tren yang dilakukan sebagai upaya untuk melestarikan budaya. Salah satu tren yang banyak dilakukan adalah menjadikan *image processing* sebagai media untuk membantu melestarikan budaya. Penggunaan *image processing* tersebut sangat beragam, misalnya dengan melakukan pengklasifikasian dan pengenalan aksara oleh Astuty (2016) dan Pasaribu (2015), pengklasifikasian batik oleh Meccasia (2015) dan Suciati (2016), dan pencarian gambar songket oleh Amirullah (2015). Salah satu bidang dalam *image processing* yang memiliki fokus terhadap karakter adalah *optical character recognition*. Bidang ini melakukan pengenalan karakter pada tulisan cetak dan tulisan tangan dengan mesin secara otomatis (Weaver, 1993).

Penelitian terhadap pengenalan atau klasifikasi karakter menjadi suatu tantangan bagi peneliti karena beragamnya aksara dan beragamnya metode untuk melakukan ekstraksi ciri atau fitur dan metode untuk klasifikasi. Diantaranya pengenalan karakter hindi menggunakan pendekatan *neural network* dilakukan oleh Yadav (2013) mendapatkan akurasi 90%, pengenalan karakter tulisan tangan arab yang dilakukan oleh Sahlol (2014) menggunakan pendekatan *neural network* mendapatkan akurasi 88%, sedangkan Rashad (2014) melakukan pengenalan karakter arab cetak dengan hasil akurasi 87% menggunakan KNN dan 98% menggunakan random forest. Rajashkararadhya (2009) juga melakukan pengenalan karakter angka tulisan tangan kannada yang merupakan bagian dari aksara yang ada di india mendapatkan hasil akurasi 97,8% menggunakan *nearest neighbor*, pengenalan karakter persian menggunakan random forest juga dilakukan oleh Zahedi (2012) dengan hasil akurasi tertinggi 87% dan pengenalan karakter urdu dilakukan oleh Naz (2015) menggunakan *neural network* mendapatkan hasil error sebesar 5.03%. Penelitian tentang pengenalan atau klasifikasi aksara nusantara juga pernah dilakukan misalnya aksara sunda dilakukan oleh Mubarok (2010) dan Mulia (2012), tulisan tangan batak dilakukan oleh Pasaribu (2015) dan aksara jawa

dilakukan oleh Astuty (2016) dan Syam (2013) mendapatkan hasil akurasi klasifikasi lebih dari 75%. Mubarak (2010) melakukan penelitian pengenalan aksara sunda secara *online* menggunakan *kohonen neural network* mendapatkan hasil akurasi paling besar 77,39% untuk guratan normal dan 76,52% untuk guratan dengan *noise* sedangkan Mulia (2012) melakukan pengenalan aksara sunda secara *offline* menggunakan ekstraksi ciri *zoning* dan klasifikasi *support vector machine* mendapatkan hasil paling besar 93,99% dan paling kecil 67,88%.

Pengklasifikasian tersebut menggunakan metode pendekatan *machine learning*. *Machine Learning* merupakan salah satu cabang dari disiplin ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang membahas mengenai pembangunan sistem yang berdasarkan pada data. Pengenalan atau pengklasifikasian tulisan tangan mempunyai tantangan tersendiri karena tulisan tangan yang sembrono dan kasar dapat menyulitkan pengklasifikasian (Zahedi & Eslami, 2012). Penggunaan *machine learning* pada *optical character recognition* mendapatkan hasil yang beragam. Pada penelitian Rashad (2014) metode *machine learning* yang diterapkan adalah *KNN* dan *random forest*, yang masing-masing mendapatkan tingkat akurasi 87% untuk *KNN* dan 94%, 96% dan 98% untuk *random forest*. John (2012) menggunakan *support vector machine* untuk melakukan klasifikasi dan mendapatkan 90.25% akurasi klasifikasi, sedangkan Desai (2010) menggunakan *Neural Network* dan mendapatkan akurasi klasifikasi sebesar 81.77%. Pasaribu (2015) mendapatkan hasil 30% sampai 50% dikarenakan ketidaktepatan memilih ekstraksi ciri yang sesuai dengan karakteristik tangan tersebut. Maka ekstraksi fitur atau ekstraksi ciri menjadi salah satu tahap yang berpengaruh terhadap hasil klasifikasi.

Penelitian ini memfokuskan pada ekstraksi fitur yang selanjutnya akan diimplementasikan pada *machine learning* untuk klasifikasi karakter aksara sunda. Fitur merupakan faktor yang berpengaruh dalam proses klasifikasi suatu gambar. Fitur yang dipilih harus mewakili ciri dari gambar tersebut, sehingga setiap gambar memiliki ciri yang unik. Jumlah fitur pun harus diminimalkan untuk mengurangi beban pada tahap klasifikasi (Yampolskiy, 2004), maka fitur harus dipilih dan diseleksi agar beban pada tahap klasifikasi dapat dikurangi dan juga dapat

mengoptimasi proses ekstraksi fitur karena tidak semua fitur digunakan dalam tahap klasifikasi. Tujuan ekstraksi fitur antara lain adalah untuk mengurangi data yang tidak perlu dari karakter dan hanya menyimpan informasi yang berguna untuk tahap selanjutnya yaitu klasifikasi (Naz et al., 2015). Penelitian ini akan berfokus pada ekstraksi fitur atau ciri dari karakter aksara sunda yang telah ditentukan kemudian memilih metode *machine learning* dengan fitur yang telah didapatkan. Ekstraksi fitur akan dilakukan dengan menggunakan metode *zoning*, *histogram* dan pendekatan struktural. Sedangkan metode *machine learning* yang akan digunakan untuk melakukan eksperimen pada penelitian ini adalah *Random Forest*, *K-Nearest Neighbor*, *Artificial Neural Network*, dan *Support Vector Machine*. Setelah dilakukan eksperimen maka akan didapatkan metode *machine learning* yang efisien dan unggul dengan menggunakan gabungan dari metode ekstraksi fitur yang telah ditentukan.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, maka dirumuskan masalah sebagai berikut:

1. Bagaimana cara mengekstraksi fitur dalam pengolahan citra menggunakan metode *zoning*, *histogram*, dan pendekatan struktural untuk klasifikasi karakter aksara sunda ?
2. Bagaimana mencari metode ekstraksi fitur yang optimal untuk mengklasifikasi karakter aksara sunda pada metode *machine learning* (*Random Forest*, *K-Nearest Neighbor*, *Artificial Neural Network*, *Support Vector Machine*)?
3. Bagaimana menentukan metode *machine learning* yang efisien dan unggul dalam pengklasifikasian karakter aksara sunda dengan fitur yang ditentukan.
4. Bagaimana cara merancang dan mengembangkan perangkat lunak untuk klasifikasi karakter aksara sunda dengan menggunakan metode *machine learning* pada lingkungan Python?

## 1.3. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, maka tujuan penelitian ini yaitu sebagai berikut:

1. Mendesain alur proses penelitian untuk mengekstraksi fitur karakter aksara sunda menggunakan metode *zoning*, *histogram*, dan pendekatan struktural yang kemudian digunakan untuk klasifikasi karakter aksara sunda.
2. Melakukan eksperimen untuk mendapatkan metode ekstraksi fitur yang optimal untuk pengklasifikasian karakter aksara sunda pada metode *machine learning* (*Random Forest*, *KNN*, *ANN*, *Support Vector Machine*).
3. Melakukan analisis dan perbandingan metode (*Random Forest*, *KNN*, *ANN*, *Support Vector Machine*) untuk klasifikasi karakter aksara sunda.
4. Melakukan implementasi perangkat lunak untuk klasifikasi karakter aksara sunda menggunakan metode *machine learning* pada lingkungan Python.

#### **1.4. Batasan Masalah**

Berikut merupakan beberapa batasan masalah dalam penelitian ini yaitu sebagai berikut:

1. Format gambar yang digunakan adalah .JPG
2. Data gambar yang digunakan adalah gambar satu karakter.
3. Latar belakang yang terdapat pada data gambar yang digunakan berwarna putih.
4. Kelas karakter yang digunakan berjumlah 30, kelas tersebut sesuai dengan ketentuan karakter aksara sunda yang telah dibakukan.
5. Metode *machine learning* yang digunakan berjumlah 4 yaitu, *Random Forest*, *K-Nearest Neighbor*, *Artificial Neural Network*, dan *Support Vector Machine*.
6. Gambar yang diambil untuk data *training* dan *test* adalah gambar yang memiliki nilai kontras yang normal dan pada posisi yang tegak.
7. Fokus penelitian adalah pada metode ekstraksi fitur dengan menggunakan metode *zoning*, *histogram* dan pendekatan struktural.
8. Dikembangkan menggunakan *Python* dengan *library* pendukung untuk membantu *machine learning* dan optimasi program.

9. Aplikasi yang dihasilkan dari implementasi metode ekstraksi fitur dan metode klasifikasi dibuat dalam bentuk *website*.

### 1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Bagi peneliti

Peneliti diharapkan mendapatkan pengetahuan baru mengenai pengenalan karakter aksara sunda. Selain itu peneliti proses penerapan metode *machine learning Random Forest, K-Nearest Neighbor, Artificial Neural Network, dan Support Vector Machine* untuk melakukan suatu klasifikasi khususnya klasifikasi karakter dan ekstraksi fitur pada pengolahan citra dengan menggunakan metode *zoning, histogram* dan pendekatan struktural.

2. Bagi pihak lain

Hasil penelitian ini membantu praktisi dan peneliti pada bidang budaya nusantara khususnya budaya sunda untuk melestarikan aksara sunda dengan membuat program yang membantu pengenalan karakter aksara sunda. Penelitian ini membantu peneliti lain untuk memilih parameter yang optimal dan lebih efisien untuk melakukan klasifikasi menggunakan metode *Random Forest, K-Nearest Neighbor, Artificial Neural Network, dan Support Vector Machine*. Selain itu penelitian ini membantu peneliti lain untuk memilih fitur yang optimal untuk digunakan dalam proses klasifikasi.

### 1.6. Struktur Organisasi Skripsi

Untuk mempermudah penulisan skripsi ini, penulis membuat sebuah sistematika penulisan yang berisi gambaran umum dan format-format skripsi. Sistematika penulisan yaitu sebagai berikut:

#### **BAB I PENDAHULUAN**

Pada bab ini berisi latar belakang penelitian, mengapa penelitian ini dilakukan. Selain itu, dijelaskan juga rumusan masalah, tujuan penelitian, batasan masalah dan manfaat penelitian. Serta terdapat struktur organisasi penulisan skripsi ini.

#### **BAB II KAJIAN PUSTAKA**

Pada bab ini dijelaskan dasar teori yang berhubungan dengan penelitian. Seperti studi pendahuluan yang berhubungan dengan penelitian ini seperti identifikasi, karakter aksara sunda, pengolahan citra, ekstraksi fitur, metode *zoning*, metode *histogram*, metode pendekatan struktural, *machine learning*, *random forest*, *K-Nearest Neighbor*, *Artificial Neural Network*, dan *Support Vector Machine*.

### **BAB III METODOLOGI PENELITIAN**

Bab ini berisi tahap-tahap yang dilakukan pada penelitian ini, yaitu alat dan bahan penelitian, desain penelitian serta metode penelitian yang akan digunakan pada penelitian ini. Bagian-bagian tersebut akan dijelaskan secara lengkap pada bab ini.

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

Pada bab ini akan dipaparkan hasil dari penelitian yang telah dilakukan. Selain itu juga akan dibahas secara lengkap tahap-tahap atau proses yang dilakukan pada penelitian ini. Yaitu tentang proses pengumpulan data, pengembangan model, implementasi sistem, studi kasus, desain eksperimen, dan analisa.

### **BAB V KESIMPULAN DAN SARAN**

Pada bab ini berisi kesimpulan dari penelitian yang telah dilakukan. Kesimpulan merupakan jawaban dari tujuan penelitian yang terdapat pada bab ini. Selain itu pada bab ini juga terdapat saran yang merupakan kumpulan saran dan rekomendasi dari penulis untuk penelitian selanjutnya.