

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Tulisan tangan masih banyak digunakan meskipun proses tulis menulis telah dapat dilakukan dengan mesin atau komputer. Karena itu, munculah penelitian mengenai pengenalan tulisan tangan agar dapat dikenali komputer. Penelitian mengenai sistem pengenalan karakter telah dilakukan sejak tahun 1923 dengan ditemukannya *Optical Character Recognition* (OCR) yang masih sangat sederhana. Namun, penelitian sistem pengenalan karakter untuk komersial, pertama kali dimulai pada tahun 1953. Kemudian pada tahun 1965, sistem pengenalan karakter mulai digunakan secara luas untuk keperluan penyortiran surat di kantor pos Amerika Serikat dan Kanada.

Sistem pengenalan karakter sangat berguna dalam beberapa aplikasi yang memerlukan pemrosesan data tulisan tangan dengan jumlah yang besar, contohnya seperti pengenalan alamat dan kode pos, interpretasi jumlah uang pada cek, analisis dokumen dan verifikasi tandatangan. Sistem pengenalan karakter juga telah mengalami kemajuan besar namun kinerjanya belum dapat menyamai kinerja manusia. Kesulitan utama disebabkan gaya tulisan tangan manusia yang beraneka ragam, kemiripan karakter satu dengan yang lainnya dan keberagaman bentuk karakter yang tidak terbatas. Selain itu, kualitas gambar teks yang buruk, kehadiran *noise* pada latar belakang gambar yang tidak dapat dihindarkan, dan distorsi menambah kesulitan proses pengenalan karakter. Meski begitu,

pengenalan karakter tulisan tangan masih terbuka lebar untuk diteliti dan dijadikan topik menarik (Jain dan Lazzerini, 2003: 1).

Sistem pengenalan karakter secara garis besar memiliki tahapan-tahapan sebagai berikut (Cheriet, *et all*, 2007: 4).

1. Praproses

Tahapan untuk meningkatkan kualitas masukan gambar dan menempatkan data yang akan diinginkan.

2. *Feature Extraction*

Tahapan untuk merekam karakteristik khusus karakter (*feature vector*) yang telah dikuantisasi untuk proses pengenalan

3. Pengelompokan

Tahapan yang memproses *feature vector* untuk menghasilkan kelompok yang dicari.

Banyak metode yang dapat digunakan untuk mengelompokan *feature vector* yang telah didapat. Beberapa metode yang umum digunakan untuk menstrukturisasi *feature* diantaranya dengan pendekatan statistik, *neural network*, algoritma genetik dan *fuzzy clustering* (Krasteva, 2002: 1). Tidak ada satupun metode yang dinobatkan sebagai metode terbaik. Hal ini tergantung banyak faktor seperti *training set*, jumlah parameter dan lain sebagainya (Vamvakas, 2007: 26).

Menurut Krasteva (2002: 1) akurasi pengenalan karakter lebih bergantung pada pemilihan *feature*. Pernyataan ini diperkuat pendapat Cheriet, *et all* (2007: 9) bahwa keakurasian ekstraksi data menentukan kinerja sistem pengenalan karakter.

Namun, kualitas hasil ekstraksi *feature* juga, bergantung pula pada kualitas data yang dipakai.

Skripsi ini menggunakan algoritma *Fuzzy C-Means* (FCM) yang merupakan salah satu metode dari *fuzzy clustering*. FCM mengelompokkan data-data menjadi beberapa cluster (kelompok) yang masing-masing diwakili pusat cluster. Pusat cluster inilah yang dijadikan dasar untuk mengenali data baru.

Penggunaan FCM untuk pengenalan karakter tulisan tangan telah dilakukan Suharyanto (2001) dalam skripsi berjudul 'Penggunaan Fuzzy Clustering dalam Pengenalan Karakter Angka'. Nilai akurasi relatif yang dihasilkan yaitu sebesar 70%. Dengan algoritma yang sama diharapkan dapat menghasilkan nilai akurasi relatif yang lebih tinggi untuk pengenalan aksara Sunda.

Aksara Sunda merupakan salah satu kajian menarik dari penelitian ini karena penelitian yang melibatkan aksara Sunda di dalamnya relatif masih sedikit dan hanya dilakukan di lingkungan terbatas.. Aksara Sunda juga merupakan salah satu aksara tradisi hasil karya ortografi masyarakat Sunda melalui perjalanan sejarahnya sejak 5 abad yang lalu hingga saat ini.

Sejak April 2008, aksara Sunda sudah resmi masuk kedalam Unicode. Unicode adalah standard industri dalam pengodean karakter yang dirancang untuk memungkinkan teks dan simbol dari semua sistem tulisan di dunia agar dapat ditampilkan dan dimanipulasi secara konsisten oleh komputer. Salah satu manfaat aksara Sunda masuk kedalam Unicode yaitu dapat menyediakan bahan

pengembangan yang lebih luas dalam khazanah komputerisasi bahasa yang akan semakin diperlukan di masa depan.

## **1.2 Perumusan Masalah**

Skripsi ini mengkaji penggunaan algoritma FCM untuk pengenalan tulisan tangan aksara Sunda Ngalagena. Hal ini dilakukan untuk mengetahui

1. Bagaimana penerapan algoritma FCM untuk pengenalan tulisan tangan aksara Sunda.
2. Bagaimana nilai akurasi relatif algoritma FCM untuk pengenalan aksara Sunda dibandingkan nilai akurasi relatif pada penelitian Suharyanto untuk pengenalan karakter angka.

## **1.3 Batasan Masalah**

Batasan masalah dari penelitian ini adalah

1. Tulisan tangan aksara Sunda yang dapat dikenali oleh sistem yaitu aksara Sunda standard yang ditulis sesuai dengan kaidah-kaidah baku.
2. Sistem hanya dapat menggunakan dataset khusus yang dibuat sendiri penulis.
3. Sistem hanya dapat mengenali karakter tunggal.
4. Penelitian ini hanya mencakup aksara Sunda Ngalagena.

## **1.4 Tujuan Penelitian**

Adapun tujuan dari skripsi ini adalah

1. Mempelajari algoritma FCM untuk pengenalan tulisan tangan aksara Sunda.
2. Membangun prototipe sistem pengenalan aksara Sunda

3. Mengevaluasi keberhasilan sistem berdasarkan nilai akurasi relatif pengenalan karakter yang dihasilkan.

### **1.5 Manfaat Penelitian**

Beberapa manfaat yang dapat diperoleh dari skripsi ini adalah sebagai berikut.

1. Mengetahui seluk beluk aksara Sunda
2. Ikut serta dalam rangka memelihara dan mengembangkan bahasa dan sastra daerah melalui teknologi informasi.
3. Dapat dikembangkan menjadi piranti yang lebih unggul guna penelitian naskah-naskah kuno dan artefak lainnya yang menggunakan aksara Sunda.

### **1.6 Sistematika Penulisan**

Sistematika penyusunan skripsi ini dibagi menjadi beberapa bab sebagai berikut:

#### **BAB I PENDAHULUAN**

Bab I berisi pendahuluan skripsi, menjabarkan mengenai latar belakang pemilihan topik skripsi, rumusan masalah yang menjadi fokus skripsi, tujuan yang ingin dicapai dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab II berisi tinjauan pustaka yang merupakan dasar teori dari skripsi ini, yaitu pendefinisian pengenalan tulisan tangan secara umum, penjabaran mengenai prinsip-prinsip algoritma FCM dan penjelasan mengenai karakter-karakter pada aksara Sunda.

### **BAB III METODOLOGI PENELITIAN**

Bab III berisi metodologi penelitian, menjabarkan mengenai teknis pelaksanaan penelitian, mulai dari alat dan bahan penelitian, desain penelitian, dan proses yang terjadi.

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

Bab IV berisi hasil penelitian dan pembahasan, menjabarkan hasil penelitian yang dilakukan beserta analisisnya.

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

Bab V berisi kesimpulan dan saran dari pengerjaan skripsi ini.

