BAB 1
PENDAHULUAN

1.1. Latar Belakang Masalah

Pengenalan citra pada huruf atau sering disebut dengan Optical Character Recognition (OCR) merupakan konversi dari tulisan tangan ke dalam bentuk file citra yang diubah.Tulisan tangan tersebut nantinya di-scan menggunakan scanner untuk menghasilkan suatu citra.[1]. Teknik OCR biasa diterapkan pada berbagai tulisan tangan, salah satunya membahas tulisan aksara Jepang.Karakter aksara Jepang itu lebih rumit jika dibandingkan dengan karakter aksara romawi, selain itu aksara jepang lebih banyak jumlahnya. Sehingga membutuhkan waktu yang tidak sedikit untuk memahaminya.

Bahasa Jepang termasuk bahasa yang memiliki bentuk bahasa yang berbeda dengan bahasa asing lainnya.Bentuk bahasa tersebut dapat diamati dari pelafalan, kosakata, gramatikal, tata bahasa, cara-cara pengungkapan, dan ragam bahasa yang digunakan. Ragam bahasa Jepang sangat dipengaruhi oleh faktor sosial budaya seperti, wilayah atau daerah, kelas sosial, perbedaan jenis kelamin, dan usia.[4].

Mahasiswa maupun individual merasa perlu mempelajari huruf Hiragana dan Katakana akan tetapi banyak kesulitan kesulitan yang dialami pada saat mempelajari bahasa jepang antara lain sulit untuk mengenali huruf satu dengan yang lainnya karena huruf katakana dan hiragana berjumlah banyak dan hamper mirip, selain itu sulit mengetahui tingkat keakuratan hasil tulisan tangan bahasa jepang karena padat tulisan tangan bahasa jepang setiap garis memiliki makna tersendiri.[5].Sehingga dibutuhkan alat bantu untuk mempermudah mengenali huruf hiragana dan katakana, salah satunya dengan pengenalan citra aksara jepang.

Banyak metode yang bisa digunakan untuk pengenalan citra aksara jepang, antara lain metode Artificial Neural Network [8], Backpropagation, Hough Transform [13], dan Support Vector Machine [10]. Dari beberapa metode yang disebutkan metode Support Vector Machine paling banyak digunakan

karenamemiliki kelebihan dalam akurasi yang baik yakni 90% [15]. Support Vector Machine (SVM) adalah suatu sistem pembelajaran yang menggunakan ruang hipotesis dari suatu fungsi linear dalam suatu ruang dimensi berfitur tinggi yang dikembangkan oleh Boser, Guyon, Vapnik.Metode *Support Vector Machine*(SVM) dalam penelitian ini digunakan karena diimplementasikan untuk berbagai macam penelitian dengan efisien dalam penggunaannya[2].

Metode *Support Vector Machine* sudah pernah digunakan dalam mengklasifikasikan beberapa kasus, seperti penerapan metode SVM pada pengenalan akasara sunda[2], penerapan metode *Support Vector Machine* pada data akreditasi sekolah SD[3], Identifikasi tanda tangan [4], pengenalan pola aksara jawi[1], pengenalan tulisan tangan angka dan operator matematika berdasarkan Algoritma *Support Vector Machine dan Hybird* [4], metode *Support Vector Machine Multiclass (One Vs All),* metode ini digunakan karena pada penelitian perbandingan *multiclass (svm)* [3], telah mendapatkan rata-rata akurasi terbaik sebesar 90%.

Metode Support Vector Machine dapat dikombinasikan dengan metode ekstrasi fitur *Zoning-Zone Centroid Zone* mampu menghasilkan hasil akan akurat dalam pengenalan tulisan, metode ini dipilih karena berdasarkan penelitian sebelumnya apabila metode klasifikasi SVM dan Zoning dikombinasikan akan lebih akurat[2].

Berdasarkan pemaparan yang telah dibahas, penelitian ini dimaksudkan untuk menerapkan metode *Support Vector Machine* dan *Zone Centroid Zone* pada pengenalan citra katakana dan hiragana.

1.2. Identifikasi Masalah

Berdasarkan penjelasan yang sudah diuraikan pada latar belakang di atas, maka identifikasi masalah dari penelitian ini yaitu belum diketahuinya tingkat keakuratan tulisan tangan metode ekstrasi ciri *Zoning (Zone Centroid Zone)* dikombinasikan dengan metode klasifikasi *Support Vector Machine (SVM) multiclass* pada aksara katakana dan hiragana.

1.3. Maksud dan Tujuan

Maksud dari penelitian ini membuat aplikasi pengenalan tulisan tangan menggunakan metode ekstrasi ciri *Zoning(Zone Centroid Zone)* dan metode klasifikasi *Support Vector Machine (SVM)* apabila diimplementasikan dalam kasus evaluasi pengenalan aksara katakana dan hiragana.

Tujuan yang ingin dicapai dalam penelitian ini yaitu mengetahui tingkat keakuratan pengenalan tulisan tangan menggunakan metode *Zoning(Zone Centroid Zone) Support Vector Machine (SVM)* pada aksara katakana dan hiragana.

1.4. Batasan Masalah

Agar penelitian yang dilakukan lebih terarah sesuai dengan tujuan yang diinginkan, makan berikut adalah batasa masalah yang digunakan :

1. Input (Data Masukan)
    a. Citra tulisan tangan hasil *scan* atau foto berformat(.jpg).
    b. Data latih diambil dari 10 orang data tulisan tangan.
    c. Karakter yang dapat dikenali hanya berupa huruf katakana dan hiragana.
    d. Kertas yang digunakan untuk menulis aksara katakana dan hiragana adalah kertas dengan background putih dan ditulis menggunakan pulpen maupun spidol berwarna hitam.
    e. Data Set dan data uji diambil dari responden mahasiswa/i sastra jepang Universitas Komputer Indonesia (UNIKOM)
2. Proses
    a. Metode *Preprocessing* yang digunakan *Grayscaling, Sauvola Threshold.*
    b. Ekstrasi Fitur yang digunakan *Zoning (Zone Centroid Zone)*
    c. Metode klasifikasi digunakan *Support Vector Machine (One vs All)*

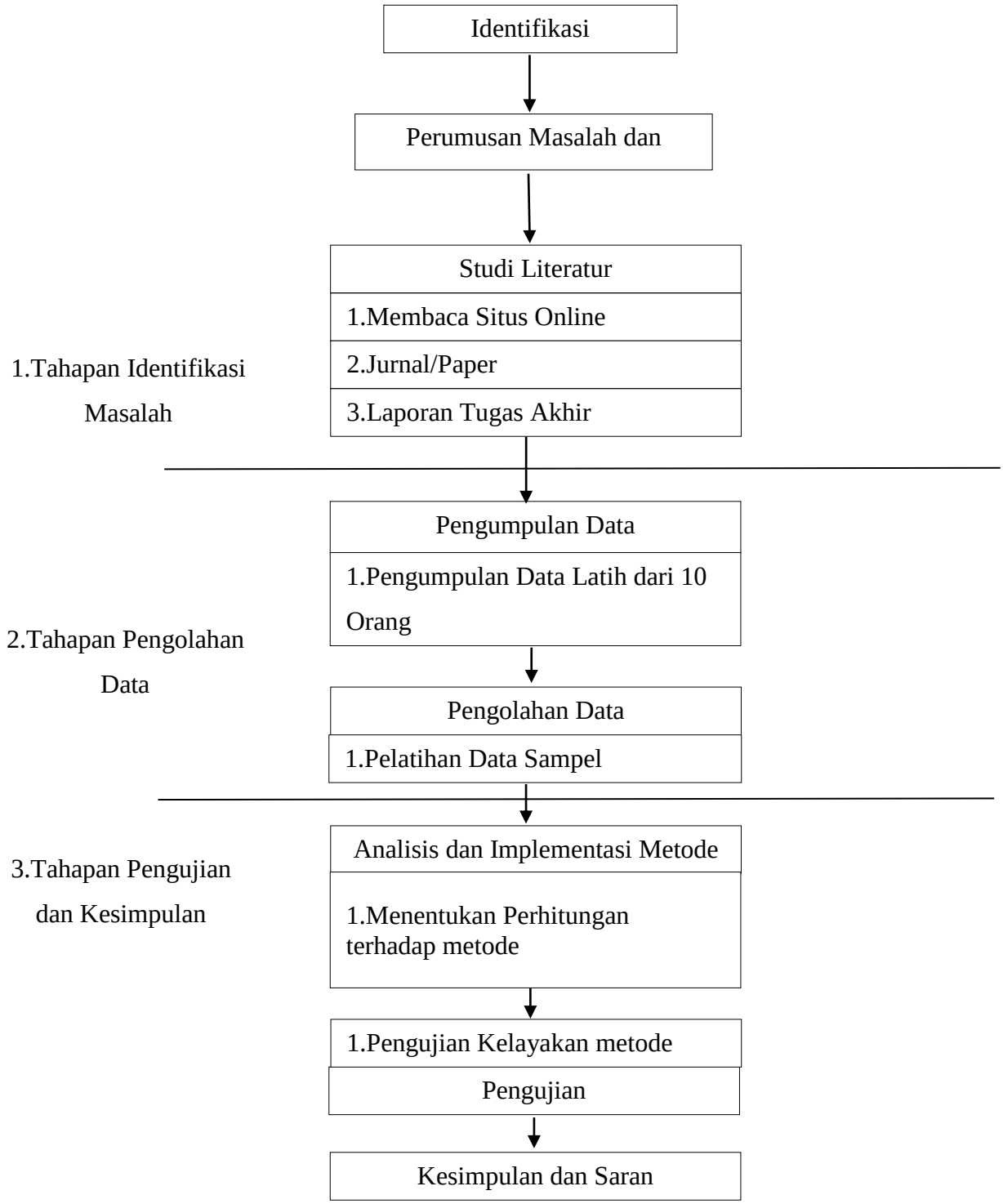
3. Output

a. Evaluasi aksara katakana dan hiragana yang dilakukan hanya sampai pada menentukan tingkat teks antara tulisan tangan dengan contoh penulisan yang benar.
b. Penilaian berdasarkan tingkat kesamaan teks tulisan tangan dengan contoh penulisan yang benar.

1.5. Metodologi Penelitian

"Metodologi penelitian pada dasarnya merupakan cara ilmiah untuk mendapatkan data yang valid dengan tujuan yang bersifat penemuan, pembuktian dan pengembangan suatu pengetahuan sehingga hasilnya dapat digunakan untuk memahami, memecahkan dan mengantisipasi masalah"[6].

Metodologi yang digunakan adalah metode SDLC (*Software/System Development Life Cycle*[7].metode penelitian ini digunakan karena sesuai dengan tahapan penelitiannya sesuai dengan yang akan dilakukan oleh peneliti.



Gambar 1. 1 SkemaPenelitian

1.6. Metode Pengumpulan Data

Metode pengumpulan data merupakan langkah yang paling strategis dalam penelitian, karena tujuan utama dari penelitian adalah mendapatkan data[7].

a. Studi Literatur

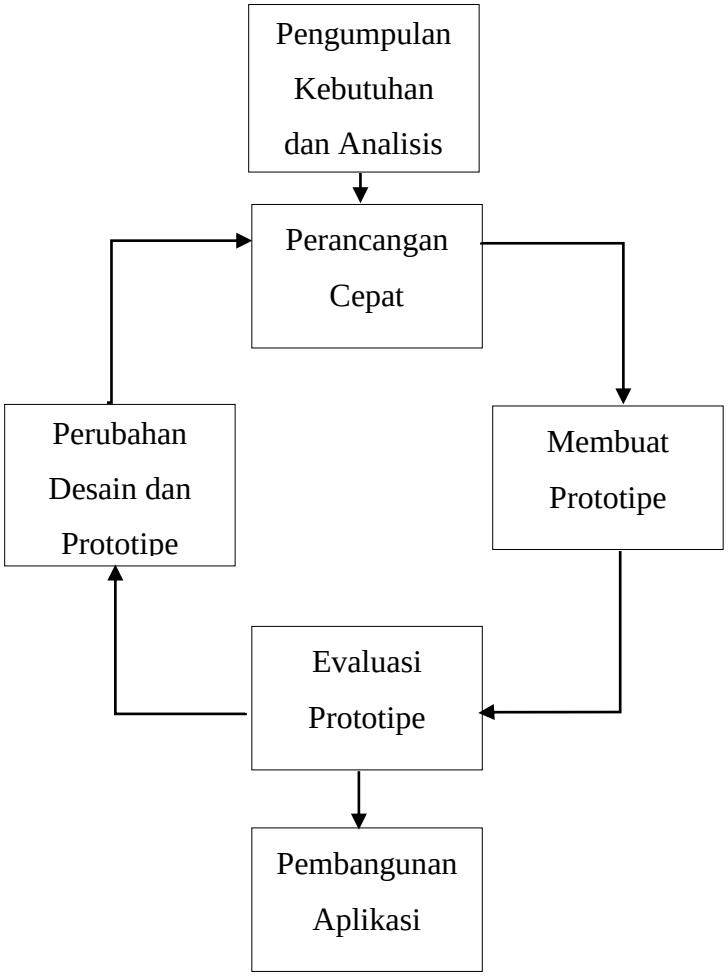
   Studi literatur adalah pengumpulan data melalui buku-buku, jurnal, *paper*, dan bacaan-bacaan yang berkaitan dengan penelitian ini.

b. Pengumpulan sampel tulisan tangan

   Pada tahapan ini peneliti mengumpulkan sampel karakter tulisan tangan dari 10 orang. Data sampel ini digunakan sebagai data training saat melakukan pelatihan dengan metode ekstrasi ciri zoning (ZCZ) dan metode klasifikasi *Support Vector Machine (SVM).*

1.7. Metode Pembangunan Perangkat Lunak

Dalam penelitian ini pembangunan perangkat lunak yang digunakan adalah model prototipe[5] karena metode ini menyajikan gambaran lengkap dari suatu sistem perangkat lunak yang akan dibangunnya, penentuan kebutuhan lebih mudah diwujudkan, waktu perancangan antara peneliti dan calon pengguna. Berikut ini adalah proses dari model prototipe pada gambar 1.2.



Gambar 1. 2 Model Prototipe

1.8. Sistematika Penulisan

Adapun sistematika penulisan laporan ini adalah sebagai berikut :

Bab 1 Pendahuluan

Bab ini berisi latar belakang, perumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian, dan sistematika penulisan.

Bab 2 Tinjauan Pustaka

Bab ini merupakan bab yang berisi tinjauan pustaka dari penelitian-penelitian sebelumnya yang telah dilakukan dan penelitian yang berkaitan dengan penelitian yang sedang dilakukan

Bab3 Analisis dan Perancangan

Bab yang berisi analisis sistem yang diperlukan dalam pengembangan dan juga rancangan sistem yang akan dibuat.

Bab 4 Implementasi dan Pengujian

Bab yang berisi penjelasan implementasi sistem yang dilakukan dalam penelitian ini dan pengujian terhadap sistem yang telah dibangun.

Bab 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dari penelitian ini berdasarkan hasil dan pembahasan serta saran-saran untuk pengembangan penelitian ini.