

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

*Optical Character Recognition* (OCR) atau disebut juga pengenalan character pada citra yang berfungsi untuk mengubah citra karakter menjadi bentuk teks. OCR juga merupakan teknik pengenalan karakter melalui mekanisme optik secara otomatis dan dapat mengenali teks hasil cetakan. Teknik OCR bisa diterapkan kedalam bentuk dokumen seperti sertifikat, ktp, sim, dan lain-lain [1].

Dalam penelitian Anggi Sanjaya Kusnanto dengan judul implementasi ocr (*optical character recognition*) menggunakan metode string matching untuk mendeteksi obat dan makanan berbasis android mempunyai akurasi 73% [2]. Kemudian dalam penelitian Faisal, Jyoti, Milan, dan Pratik dengan judul implementasi OCR menggunakan *Pattern Matching* pada gambar karakter dari huruf yang sama tingkat akurasinya besar. Namun, akurasi akan menurun menjadi 70% ketika orang-orang tidak menggunakan huruf yang sama pada saat pengujian dengan data latihnya[3]. Kemudian menurut Reza Yogi Andria pada judul OCR menggunakan Support Vector Machine (SVM) dan Zoning Pada sertifikat karakter pada citra sertifikat tidak semua sertifikat memiliki jenis huruf yang sama seperti dokumen lainnya melainkan berbeda jenis huruf pada setiap sertifikat[4].

Sertifikat merupakan surat keterangan pernyataan secara tertulis atau tercetak dari orang yang berwenang yang dapat digunakan sebagai bukti suatu kejadian [5]. Sertifikat harus di simpan dengan baik agar bisa di pergunakan seperti mengajukan lamaran kerja, persyaratan skripsi, sertifikat tanah, dan lainnya. Setiap sertifikat memiliki bagian atau informasi seperti nomor surat, nama peserta, waktu dan tempat, serta pengesahan pada sebuah sertifikat, pada sebuah sertifikat terdapat jenis jenis huruf karena jenis huruf yang ada di sertifikat tidak selalu sama dan juga berbeda-beda. Untuk itu objek yang akan digunakan sebagai bahan penelitian adalah hasil *scan* citra sertifikat.

Akurasi teks yang dikenali oleh OCR bergantung pada metode pengenalan yang digunakan. Salah satu metode yang dapat digunakan dalam pengenalan karakter pada sertifikat adalah menggunakan metode *Learning Vector Quantization* (LVQ). LVQ adalah suatu metode jaringan syaraf tiruan (JST) yang melakukan pembelajaran pada lapisan kompetitif yang terawasi (*supervised learning*) yang arsitekturnya bersifat tunggal (*single layer*)[6]. Penelitian yang terkait dengan metode LVQ yang dilakukan oleh Lilik Anifah dibidang klasifikasi karakter dengan akurasi algoritma LVQ mampu bekerja pada rata-rata tingkat akurasi sebesar 84,43%[7]. Sedangkan penelitian yang dilakukan Maharani Dessy Wuryandari dan Irawan Afrianto pada dalam membandingkan metode Jaringan Syarat Tiruan *Backpropagation* dan *Learning Vector Quantization* pada pengenalan wajah klasifikasi LVQ mendapatkan akurasi sebesar 37,63% sedangkan pada metode klasifikasi BP mendapatkan akurasi sebesar 37,33% [8].

Oleh karena itu, Algoritma klasifikasi LVQ memiliki memungkinkan untuk memperoleh akurasi yang lebih baik apabila dilakukan pada pengenalan karakter dokumen sertifikat. Karena metode LVQ memiliki beberapa keuntungan utama yaitu kinerja jaringan yang baik, pelatihannya cepat, jumlah neuron lebih sedikit, tingkat pengenalan tinggi, struktur jaringan sederhana, dan tidak memerlukan vektor masukan dinormalisasi [9].

## **1.2 Rumusan Masalah**

Berdasarkan penjelasan yang sudah diuraikan pada latar belakang, maka identifikasi masalah dari penelitian ini yaitu sulitnya mengenali huruf dan angka karena pada setiap dokumen sertifikat terdapat jenis huruf dan ukuran yang berbeda-beda.

## **1.3 Maksud dan Tujuan**

Maksud dari penelitian ini adalah melakukan implementasi metode klasifikasi *Learning Vector Quantization* (LVQ) pada kasus dokumen sertifikat. Tujuan yang ingin dicapai dalam penelitian ini yaitu untuk mengetahui tingkat keakuratan metode klasifikasi *Learning Vector Quantization* (LVQ) dalam pengenalan karakter pada dokumen sertifikat.

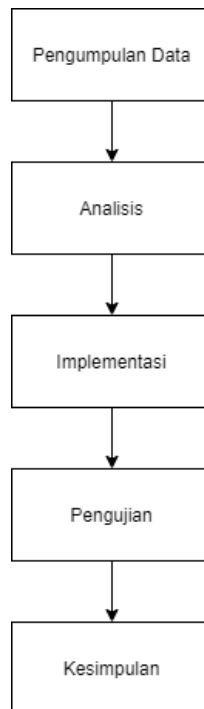
#### 1.4 Batasan masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data masukan yang digunakan berupa format .jpg dan .png.
2. Data masukan diperoleh dari umum dan internet.
3. Dokumen sertifikat harus memiliki kualitas gambar yang baik agar dapat terdeteksi pada preprocessing dan segmentasi
4. Data yang dilakukan pengenalan hanya huruf dan angka, untuk simbol dan *special character* tidak dilakukan proses pengenalan.
5. Data yang dikumpulkan berupa *dataset* huruf (A - Z), (a - z), dan angka (0 - 9).
6. Karakter yang akan diambil untuk pengenalan adalah bagian tengah pada citra sertifikat.
7. Sertifikat yang memiliki background dapat mempengaruhi dalam proses pengenalan karakter.

#### 1.5 Metode Penelitian

Metode yang digunakan pada penelitian ini memiliki beberapa tahap, yaitu tahap pengumpulan data, analisis, implementasi, pengujian, dan kesimpulan. Metode pada penelitian ini dapat dilihat pada gambar 1.1 dibawah ini:



**Gambar 1.1 Metode Penelitian**

### 1.5.1 Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini memiliki beberapa tahapan yaitu sebagai berikut:

a. Studi Literatur

dilakukan dengan cara mengumpulkan data dan informasi dari literatur, jurnal, paper, buku, serta bacaan lainnya yang berkaitan dengan *Optical Character Recognition* dan *Algoritma Learning Vector Quantization* yang akan digunakan dalam penelitian.

b. Pengumpulan Dataset

Dataset yang akan digunakan dalam penelitian ini adalah berupa citra karakter yang terdiri dari 62 huruf diantaranya [A-Z] [a-z] [0-9] berdasarkan beberapa jenis font pada data latih yang didapatkan dari *paint* atau *photoshop*, sedangkan untuk data uji menggunakan hasil scan citra sertifikat. Dataset yang digunakan sebanyak 10 citra font dengan total citra sebanyak 12.400 atau 200 citra per karakter dan 4 citra uji hasil scan

### 1.5.2 Analisis

Analisis yang dilakukan pada penelitian ini dibagi menjadi dua yaitu:

1. Analisis Data Masukan

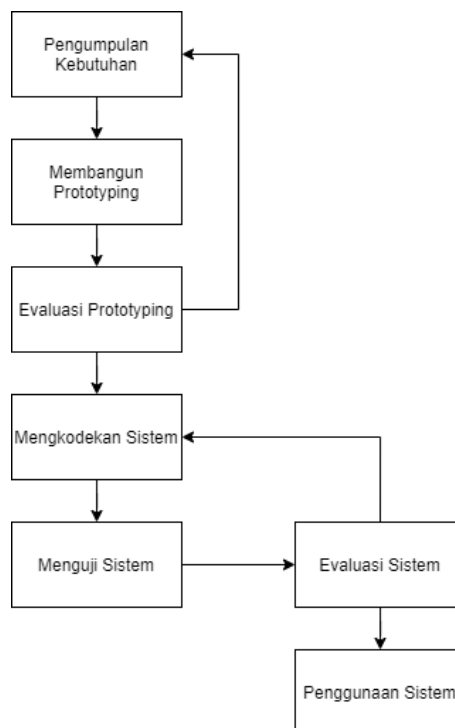
Analisis ini bertujuan untuk menganalisis data masukan yang akan diproses, yaitu berupa karakter-karakter pada sebuah jenis huruf tertentu.

2. Analisis Kebutuhan Pembangunan

Analisis ini bertujuan untuk menganalisis kebutuhan seperti metode dalam preprocessing dan LVQ.

### 1.5.3 Implementasi

Metode yang akan digunakan dalam penelitian ini adalah metode Prototype. Metode prototype digunakan karena pada metode ini dapat melakukan penelitian secara berulang-ulang, atau dapat terus di perbaharui (*update*) [10]. Implementasi dapat dilihat pada gambar 1.2 berikut.



**Gambar 1.2 Model Prototype**

1. Pengumpulan Kebutuhan

Pengumpulan kebutuhan dilakukan dengan cara melakukan penelitian dengan menggunakan metode pengumpulan data sehingga memperoleh kebutuhan yang dibutuhkan

2. Membangun *Prototyping*

Membangun *prototyping* adalah membuat perancangan sementara yang berfokus pada output yang akan di tampilkan

3. Evaluasi *Prototyping*

Evaluasi *Prototyping* adalah melakukan evaluasi sesuai dengan kebutuhan sistem yang akan dibangun. Jika sesuai maka Langkah 4 akan diambil, jika tidak sesuai maka mengulangi Langkah 1, 2, dan 3.

4. Mengkodekan Sistem

Dalam tahap ini *prototyping* yang sudah disepakati diterjemahkan kedalam Bahasa pemrograman yang telah direncanakan.

5. Menguji Sistem

Setelah sistem sudah menjadi perangkat lunak yang siap di pakai, maka harus diuji dengan beberapa pengujian seperti, Black Box, White box, Pengujian arsitektur, Basis path dan lain-lain.

6. Evaluasi Sistem

Di tahap ini klien mengevaluasi sistem yang sudah dibuat sudah sesuai yang diinginkan. Jika tidak, maka pengembang akan mengulangi langkah ke 4 dan 5. Tapi jika iya, maka langkah ke 7 akan dilakukan.

7. Penggunaan Sistem

Perangkat lunak yang telah diuji dapat digunakan.

#### **1.5.4 Pengujian**

Pada tahap ini akan dilakukan pengujian akurasi dari metode *Learning Vector Quantization* (LVQ) terhadap aplikasi *Optical Character Recognition* (OCR) akan dihitung seberapa besar hasil akurasi yang diperoleh dari pengimplementasian metode LVQ pada kasus ekstraksi dokumen citra sertifikat.

### 1.5.5 Penarikan Kesimpulan

Pada tahap ini, akan menarik kesimpulan berdasarkan pengujian terhadap penerapan metode klasifikasi *Learning Vector Quantization* (LVQ) dalam mengenali teks pada sertifikat.

### 1.6 Sistematika Penulisan

Untuk menentukan gambaran umum mengenai penelitian yang akan dilakukan oleh penulis, terdiri dari lima bab dengan sistematika penulisan sebagai berikut:

#### **BAB 1 PENDAHULUAN**

Bab ini menguraikan tentang dasar-dasar pemikiran yang berisi tentang latar belakang masalah, rumusan masalah, maksud dan tujuan penelitian, batasan masalah, metode penelitian, serta sistematika penulisan.

#### **BAB 2 DASAR TEORI**

Bab ini membahas mengenai konsep dasar dan teori yang berkaitan dengan penelitian yang dilakukan yang berguna dalam hal proses analisis permasalahan serta tinjauan terhadap penelitian - penelitian serupa yang telah pernah dilakukan sebelumnya. Teori tersebut antara lain seperti OCR, *Preprocessing*, MSER, Segmentasi, scaling, ekstraksi ciri dan algoritma LVQ

#### **BAB 3 ANALISIS DAN PERANCANGAN SISTEM**

Bab ini menjelaskan mengenai cara kerja sistem mengidentifikasi masalah dan menjalankan proses untuk menyelesaikan masalah tersebut, seperti analisis sistem, analisis masalah, analisis metode, analisis kebutuhan non fungsional, pemodelan sistem, serta perancangan antar muka.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini menjelaskan mengenai implementasi dan pengujian sistem. Tahap implementasi merupakan tahap pengimplementasian sistem untuk menguji sistem yang telah dibangun. Metode yang akan

digunakan untuk melakukan pengujian sistem adalah menggunakan metode *Black-box* serta pengujian untuk mengetahui akurasi dari metode *Learning Vector Quantization (LVQ)* dengan cara menghitung akurasi dari hasil klasifikasi sistem tersebut.

## **BAB 5 KESIMPULAN DAN SARAN**

Bab ini menjelaskan mengenai kesimpulan yang diperoleh dari pembuatan Tugas Akhir ini, serta saran-saran yang akan dijadikan pengembangan dimasa yang akan datang.