

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Optical Character Recognition (OCR) adalah suatu perangkat lunak yang dapat mengidentifikasi citra huruf ataupun angka yang nantinya dapat di konversikan dalam bentuk dokumen digital atau file text[1]. Dengan OCR yang memiliki prinsip seperti meniru manusia dalam membaca, memindai sesuatu objek yang didalamnya terdapat teks, serta menginterpretasikan teks yang terdapat didalam objek tersebut, membuat OCR menjadi solusi yang cukup efektif untuk proses konversi dokumen cetak kedalam bentuk dokumen digital[2].

Pada penelitian terhadap penggunaan Teknologi OCR dengan memanfaatkan library bawaan Matlab untuk mengekstraksi teks rambu lalu lintas, didapatkan hasil uji tingkat kesuksesan ekstraksi teks mencapai 97% serta kesesuaian hasil ekstraksi teks dengan kecocokan informasi pada rambu petunjuk arah sebesar 60%[3]. Sedangkan pada penelitian lain terhadap OCR menggunakan Algoritma *Template Matching* pada kasus pengenalan citra digital berisi teks didapatkan akurasi rata-rata tingkat keberhasilan mencapai 92,90%, Namun akurasi dapat turun apabila terdapat karakter yang tidak terdaftar pada *Template Matching*[1].

Sertifikat menurut Kamus Besar Bahasa Indonesia (KBBI) merupakan Kata Nomina atau kata benda. Dengan kata lain, sertifikat adalah tanda ataupun surat pernyataan yang berbentuk tertulis maupun tercetak yang diberikan langsung oleh pihak yang berwenang dan dapat digunakan sebagai bukti dari kepemilikan atau suatu kejadian[4]. Namun pada sertifikat, jenis karakter yang dimiliki cenderung tidak sama dengan jenis karakter pada dokumen biasanya, serta setiap sertifikat pun memiliki tulisan yang berhimpit[5]. Untuk itu objek yang akan digunakan sebagai bahan penelitian adalah hasil scan karakter yang terdapat pada sertifikat

Berdasarkan penelitian terhadap Pengenalan Aksara Sunda yang memanfaatkan metode *Learning Vector Quantization* (LVQ) sebagai metode klasifikasinya, didapatkan bahwa akurasi pengenalan aksara mencapai 78,67%, akurasi tersebut berpegang pada data latih, ukuran aksara, dan teknik pengambilan citra aksara[6]. Pada penelitian lain yang menggunakan metode klasifikasi *Learning Vector Quantization* (LVQ) pada kasus identifikasi nomor kendaraan secara otomatis, didapatkan bahwa akurasi pengenalan mencapai 95,32%[7]. Sedangkan pada penelitian lain yang menggunakan *Learning Vector Quantization 3* (LVQ3) sebagai metode klasifikasi pada kasus klasifikasi potensi banjir di kabupaten malawi berbasis web, yang dimana LVQ 3 merupakan kembangan dari LVQ 1, didapatkan bahwa hasil akurasi pelatihan paling besar mencapai 97,62%[8]. Berdasarkan hal tersebut, pada penelitian ini OCR akan dikenali dengan menggunakan metode klasifikasi *Learning Vector Quantization 3* (LVQ3) untuk mendapatkan akurasi yang lebih baik.

1.2 Rumusan Masalah

Berdasarkan dari latar belakang diatas, maka dapat disimpulkan sebuah perumusan masalah yaitu apakah dengan menerapkan metode klasifikasi *Learning Vector Quantization 3* (LVQ3) dapat memperoleh akurasi yang baik pada pengenalan karakter citra sertifikat

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk mengimplementasikan metode *Learning Vector Quantization 3* untuk klasifikasi pada kasus pengenalan karakter citra sertifikat. Sedangkan tujuan dari penelitian ini adalah untuk mengetahui hasil akurasi pengenalan karakter yang diperoleh setelah menerapkan metode *Learning Vector Quantization 3*

1.4 Batasan Masalah

Batasan-batasan masalah yang terdapat pada penelitian ini adalah :

1. Penginputan berupa citra baik citra sertifikat maupun citra karakter yang beformat .png atau .jpg

2. Dataset yang akan digunakan adalah citra teks yang terdiri dari kumpulan karakter [A-Z][a-z][0-9].
3. Hasil scan sertifikat digunakan sebagai objek untuk penelitian.
4. Karakter berupa symbol tidak dijadikan sebagai objek penelitian.
5. Sertifikat yang memiliki background dapat mempengaruhi dalam proses pengenalan karakter.
6. Karakter yang berhimpit dapat dikenali menjadi 1 karakter, sehingga dapat menurunkan akurasi pengenalan.
7. Karakter yang akan diambil untuk pengenalan adalah bagian tengah pada citra sertifikat.

1.5 Metode Penelitian

Terdapat tahap-tahap dari metode penelitian yang akan digunakan yaitu tahap pengumpulan data, tahap analisis, tahap pembangunan perangkat lunak, tahap pengujian, tahap kesimpulan :

1.5.1 Tahap Pengumpulan Data

Tahap pengumpulan data pada penelitian ini memiliki beberapa tahapan, sebagai berikut:

a. Identifikasi Masalah

Identifikasi masalah diperlukan untuk mencari tahu beberapa penelitian yang melakukan penelitian terhadap kasus pengenalan citra karakter yang nantinya akan digunakan sebagai bahan acuan pada penelitian ini.

b. Studi Literatur

Studi literatur dilakukan untuk mendapatkan pengetahuan serta referensi mengenai Optical Character Recognition maupun metode *Learning Vector Quantization 3* dengan cara mengumpulkan informasi melalui internet, jurnal, buku, dan paper.

c. 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 jenis font pada data

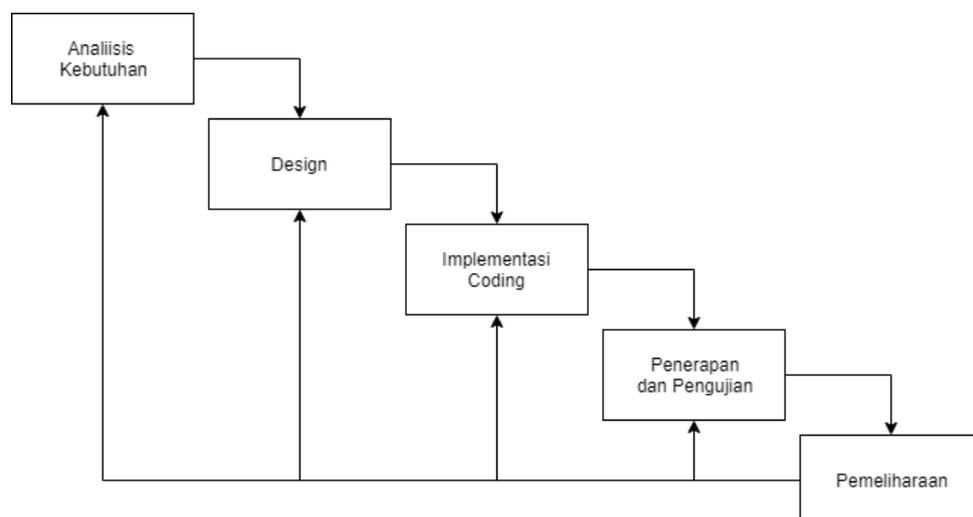
latih, sedangkan untuk data uji menggunakan hasil scan karakter yang terdapat pada citra sertifikat

1.5.2 Tahap Analisis

Tahap analisis yang akan digunakan pada penelitian ini diantaranya: Analisis Masalah, Analisis *Preprocessing*, Analisis Data Masukan, Analisis Metode *Learning Vector Quantization* 3.

1.5.3 Tahap Pembangunan Perangkat Lunak

Metode yang digunakan untuk tahapan membangun perangkat lunak pada penelitian ini adalah Metode *Waterfall*. *Waterfall* merupakan salah satu model *Software Development Life Cycle* (SDLC) yang sering digunakan atau sering juga diebut dengan model konvensional atau classic life cycle, pendekatan model ini lebih sistemasi dan berurut dimulai dari analisis kebutuhan sistem lalu menuju ketahapan analisis kebutuhan untuk menganalisa apa saja kebutuhan dari sistem, desain sistem, proses coding, testing/verification, dan terakhir maintenance atau pemeliharaan[9]. Urutan dari tahapan-tahapan pada Model *Waterfall* akan terlihat pada Gambar 1



Gambar 1. 1 Metode Waterfall

Berikut ini adalah penjelasan dari setiap tahap yang terdapat pada Metode Waterfall:

1. Analisis Kebutuhan

Pada tahap ini dilakukan analisis sebelum memulai pembangunan perangkat lunak yaitu mengetahui kebutuhan apa saja yang diperlukan didalam pembangunan perangkat lunak pada penelitian ini, hal tersebut seperti data-data yang diperlukan, fitur dan fungsi pada perangkat lunak. Dalam tahap ini kebutuhan metode yang akan digunakan pada penelitian ini adalah Learning Vector Quantization 3

2. Desain

Pada tahap ini akan dilakukan perancangan desain yang nantinya akan diterapkan pada perangkat lunak yang terdapat pada penelitian ini. Perancangan desain berupa Unified Modelling Language (UML) untuk merancang kerangka yang terdapat pada perangkat lunak dan juga perancangan antarmuka guna mengetahui tampilan seperti apa yang akan digunakan pada perangkat lunak yang dibangun.

3. Implementasi Coding

Pada tahap ini desain yang telah dirancangan akan diterapkan atau dibentuk kedalam coding sehingga menjadi perangkat lunak yang siap digunakan nantinya.

4. Penerapan dan Pengujian

Pada tahap ini perangkat lunak yang telah dibangun selanjutnya akan dicoba dan diterapkan untuk menguji fungsi-fungsi yang terdapat pada program apakah dapat berjalan dengan semestinya,

5. Pemeliharaan

Pada tahap ini akan dilakukan pemeliharaan atau maintenance pada program jika ditemukannya error atau beberapa fungsi yang tidak bekerja dengan cara memperbaikinya untuk dapat meningkatkan kinerja dari sistem yang dibangun. Tahapan ini diperlukan jika pada tahapan-tahapan sebelumnya tidak ditemukannya masalah.

1.5.4 Tahap Pengujian

Pada tahap pengujian ini, nantinya akan dihitung seberapa besar hasil akurasi yang diperoleh dari pengimplementasian metode *Learning Vector Quantization 3* pada pengenalan karakter citra sertifikat.

1.5.5 Tahap kesimpulan

Pada tahap ini, akan ditarik kesimpulan berdasarkan hasil pengujian terhadap penerapan metode klasifikasi *Learning Vector Quantization 3* dalam mengenali karakter pada citra sertifikat.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan skripsi ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Menguraikan tentang latar belakang permasalahan, mencoba merumuskan inti permasalahan yang dihadapi, menentukan tujuan dan kegunaan penelitian, yang kemudian diikuti dengan pembatasan masalah, asumsi, serta sistematika penulisan

BAB 2. LANDASAN TEORI

Membahas tempat studi kasus kemudian berbagai konsep dasar dan teori - teori yang berkaitan dengan topik penelitian yang dilakukan dan hal hal yang berguna dalam proses analisis permasalahan serta tinjauan terhadap penelitian penelitian serupa yang telah pernah dilakukan sebelumnya. Teori-teori tersebut terdiri dari teori OCR, preprocessing, image processing, sertifikat, citra, teknik klasifikasi

BAB 3. ANALISIS DAN PERANCANGAN SISTEM

Bab 3 terbagi menjadi dua bagian yaitu analisis dan perancangan sistem. Bagian analisis sistem yaitu deskripsi masalah, analisis kasus, analisis masalah, analisis kebutuhan, menguraikan cara pemecahan masalah, selain itu terdapat perancangan sistem yang terbagi menjadi perancangan komponen, perancangan berorientasi objek, dan perancangan antarmuka untuk aplikasi yang akan dibangun sesuai dengan hasil analisis yang telah dibuat.

BAB 4. IMPLEMENTASI DAN PENGUJIAN SISTEM

Menjelaskan implementasi dari hasil analisis dan perancangan yang telah dibuat ke dalam bentuk aplikasi pemograman, seperti implementais terhadap sistem ocr, pengklasifikasian citra dokumen, serta mengujian terhadap sistem yang telah dibuat apakah berjalan secara lancer atau tidak. Dan terakhir adalah pengujian akurasi terhadap ocr dalam membaca karakter yang terdapat pada sertifikat

BAB 5. KESIMPULAN DAN SARAN

Berisi kesimpulan yang sudah diperoleh dari hasil penulisan tugas akhir ini dan saran untuk pengembangan penelitian lebih lanjut.