

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Optical Character Recognition (OCR) adalah metode pengenalan karakter huruf dan angka pada citra atau gambar menjadi tulisan, citra atau gambar tersebut dapat berasal dari tulisan tangan atau dari sebuah mesin pencetak[1]. Dari proses OCR menghasilkan teks yang sesuai dengan gambar *output scanner*. Karakter atau pola yang dikenali terdapat pada sebuah gambar atau citra digital, dimana gambar huruf yang dikenali dapat berupa hasil scan, foto, print-screen, dan lain sebagainya. Sistem OCR sekarang dikembangkan bukan hanya untuk mengenali satu karakter saja, namun lebih dari itu teknik OCR sekarang digunakan untuk pengenalan teks yang didalamnya terdapat beberapa huruf, huruf – huruf tersebut dikenali satu – persatu sehingga menghasilkan sebuah teks yang sesuai dengan gambar output scanner dimana tingkat keakuratan penerjemahan karakter tergantung dari tingkat kejelasan gambar dan metode yang digunakan[2].

Penelitian mengenai *Optical Character Recognition* telah dilakukan sebelumnya oleh banyak peneliti dengan berbagai macam objek penelitian, pada penelitian ini akan menggunakan objek penelitian yaitu Sertifikat. Sertifikat adalah tanda atau surat keterangan (pernyataan) tertulis dari orang yg berwenang yang dapat digunakan sebagai bukti pemilikan atau suatu kejadian. Sertifikat pula merupakan sebuah surat sebagai tanda pengakuan bahwa seseorang telah menguasai kompetensi tertentu atau telah mengikuti acara tertentu [3]. Dalam sertifikat terdapat beberapa bagian, bagian-bagian tersebut meliputi kop atau logo organisasi, nomor surat, nama peserta, waktu dan tempat, serta pengesahan. Pada sebuah sertifikat, jenis huruf yang ada bukan jenis huruf yang biasa dipakai pada citra biasanya yang hanya menggunakan satu jenis huruf melainkan kombinasi dari beberapa jenis dan ukuran *font* yang berbeda-beda beserta dengan tipe teks nya seperti *bold*, *italic*, ataupun *underline*. Maka dari itu, objek yang akan dijadikan bahan penelitian adalah hasil scan dokumen sertifikat untuk proses pengenalan karakternya.

Pada penelitian sebelumnya Reza Yogi Andria dengan judul *Optical Character Recognition* menggunakan algoritma *Support Vector Machine (SVM)* dan *Zoning* pada Sertifikat (2019), didapatkan hasil akurasi terbaik yaitu sebesar 83,28%, yang dimana akurasi yang didapat dipengaruhi oleh *preprocessing* serta data uji yang digunakan [4].

Penelitian Jianhong Xie dengan judul *Optical Character Recognition Based on Least Square Support Vector Machine* (2009). Hasil yang didapat pada karakter sulit yang sudah dicetak, system dapat mensegmentasi karakter tersebut menggunakan *thresholding* dinamis dan hasilnya cukup bagus karena menampilkan gambar yang sesuai[5]. Penelitian Gilang Rachman Perdana, Dkk dengan judul Analisis dan Implementai *Optical Character Recognition* Menggunakan *Modified Direction Feature* dan *Least Square Support Vector Machine* (2013). Akurasi terbaik yang dihasilkan adalah sebesar 68.37%[6].

Least Square Support Vector Machine (LS-SVM) merupakan salah satu dari modifikasi SVM konvensional. Jika SVM dikarakteristik oleh permasalahan *quadratic programming* dengan pembatas berupa pertidaksamaan, maka LS-SVM sebaliknya, diformulasikan dengan menggunakan pembatas berupa persamaan. Sehingga solusi LS-SVM dihasilkan dengan menyelesaikan persamaan linear. Hal yang berbeda dengan penyelesaian SVM yang solusinya didapatkan melalui penyelesaian *quadratic programming*. Saat ini, LS-SVM banyak dilakukan pada klasifikasi dan estimasi [7].

Berdasarkan uraian diatas, penelitian ini akan membangun sebuah sistem *Optical Character Recognition* pada Sertifikat dengan menggunakan metode *Least Square Support Vector Machine* dengan mengangkat kasus pada citra Sertifikat.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang diuraikan diatas, masalah yang dapat teridentifikasi yaitu sebagai berikut :

1. Terdapat karakter huruf dan angka yang masih belum bisa dikenali dikarenakan keberagaman jenis dan ukuran *font* yang berbeda-beda beserta tipe teks yang ada pada dokumen sertifikat.

2. Belum diketahuinya tingkat akurasi dari penggunaan algoritma LS-SVM dalam mengenali karakter huruf dan angka pada kasus dokumen sertifikat.

1.3 Maksud dan Tujuan Penelitian

Berdasarkan masalah yang diteliti, maka dapat disusun maksud dan tujuan sebagai berikut :

1.3.1 Maksud

Berdasarkan masalah yang telah diuraikan, maka dapat disusun maksud dari penelitian ini adalah melakukan analisis dan implementasi metode klasifikasi *Least Square Support Vector Machine* (LS-SVM) dalam pengenalan sebuah karakter pada dokumen sertifikat.

1.3.2 Tujuan

Berdasarkan masalah yang telah diuraikan, maka dapat disusun tujuan dari penelitian ini adalah sebagai berikut :

1. Mengenali karakter huruf dan angka pada sertifikat yang memiliki keberagaman jenis dan ukuran *font* dan tipe teks yang ada.
2. Mengukur titik akurasi dari penggunaan algoritma *Least Square Support Vector Machine* untuk pengenalan karakter huruf dan angka pada dokumen sertifikat.

1.4 Batasan Masalah

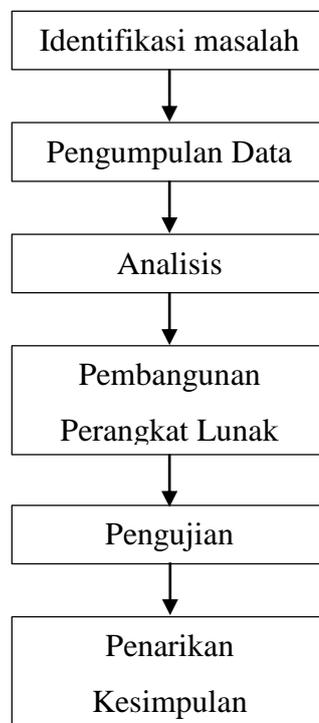
Batasan masalah pada penelitian yang akan dilakukan adalah:

1. Penelitian ini hanya melakukan pengambilan data berupa karakter angka dan huruf pada Sertifikat, untuk symbol (.,:;/[]%*'()?!#+-=) tidak akan dilakukan pengambilan karakter.
2. Sertifikat yan digunakan adalah sertifikat pada sebuah acara (seminar atau workshop).
3. Metode atau algoritma yang digunakan untuk proses klasifikasi adalah *Least Square Support Vector Machines*.
4. Tanda tangan, logo dan background dan noise tidak dilakukan proses pengenalan.

5. Jenis citra yang digunakan untuk data latih pada penelitian ini yaitu kumpulan *font* yang berformat *.jpg atau *.png.
6. Jenis citra yg digunakan untuk data test pada penelitian yaitu citra sertifikat ini berformat *.jpg atau *.png.

1.5 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian deskriptif. Metode yang mempelajari masalah yang ada dengan tatacara kerja yang berlaku. Langkah-langkah penelitian dapat dilihat pada Gambar 1.1.



Gambar 1.1 Sistematika Metode Penelitian Deskriptif

1.5.1 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan, permasalahan yang dapat diidentifikasi adalah bagaimana mengidentifikasi huruf dan angka pada sertifikat dengan menggunakan algoritma LS-SVM untuk proses klasifikasinya.

1.5.2 Pengumpulan Data

Pengumpulan data dilakukan dengan cara mengambil data dari internet, jurnal, artikel ilmiah, ataupun dari sumber-sumber yang memiliki data-data yang akan dijadikan bahan dalam penelitian yang akan dilakukan.

Kerangka Pengumpulan data yang didapatkan dari objek penelitian yaitu :

1. Studi Pustaka

Studi yang dilakukan dalam mencari informasi yang terkait dengan penelitian melalui literatur jurnal, *paper*, *proceedings*, maupun buku.

2. Pengumpulan Dataset

Pada tahap ini diambil data latih dan juga data uji. Data latih yaitu berisi kumpulan citra *font* yang berjumlah 100 citra yang terdiri dari huruf A-Z, a-z angka 0-9, untuk karakter tambahan berupa simbol (.,:;/[]%*')?!#+-=) tidak akan dimasukkan kedalam data latih. Dan untuk data test adalah hasil *scan* citra sertifikat yang akan sebanyak 10 citra sertifikat.

1.5.3 Analisis

Analisis metode adalah proses untuk menganalisa metode yang akan digunakan, metode yang digunakan dalam pembangunan aplikasi ini terbagi menjadi dua, yaitu:

- a. ***Preprocessing***

Preprocessing adalah tahapan pengolahan untuk menghilangkan bagian-bagian yang tidak diperlukan untuk diproses pada tahap Pengujian. Adapun tahapan di dalam yang ada pada tahap *preprocessing* yaitu:

1. *Grayscale*
2. *Thresholding*
3. *Resize*

- b. **Segmentasi**

Segmentasi yang digunakan pada penelitian ini adalah segmentasi *Profile Projection*

- c. **Ekstraksi Ciri**

Pada penelitian ini, metode yang digunakan untuk segmentasi adalah metode *algoritma Zone Based Feature Extraction (zoning)*.

d. Klasifikasi

Algoritma yang dipakai untuk proses klasifikasi pada sertifikat adalah algoritma *Least Square Support Vector Machine*. Pada LS-SVM proses yang akan diklasifikasi yaitu *knowledge*, *knowledge* yang dimaksud dalam hal ini adalah model hasil learning metode classifier LS-SVM yang didapat pada proses training, model inilah yang akan menjadi panduan dasar untuk proses pengujian. Baik tidaknya model classifier yang didapat selain bergantung pada kehandalan vektor ciri yang dihasilkan proses sebelumnya tetapi juga bergantung pada inisialisasi parameter LS-SVM seperti trik kernel yang dilakukan.

1.5.4 Metode Pembangunan Perangkat Lunak

Metode yang digunakan dalam pengembangan perangkat lunak adalah metode *prototype*, yang meliputi beberapa proses antara lain sebagai berikut.

a. Analisis Kebutuhan

Pada tahap ini dilakukan analisis kebutuhan aplikasi yang diperlukan dalam pembangunan aplikasi *Optical Character Recognition* menggunakan *Least Square Support Vector Machine*. Analisis dilakukan terhadap data masukan, tahapan proses, perangkat lunak, dan perangkat keras yang dibutuhkan.

b. Desain *Prototype*

Tahapan ini merupakan penjabaran dari proses sebelumnya, pada tahap ini akan ditentukan seperti apa aplikasi yang nantinya akan dibangun, metode yang digunakan dalam pembangunan aplikasi.

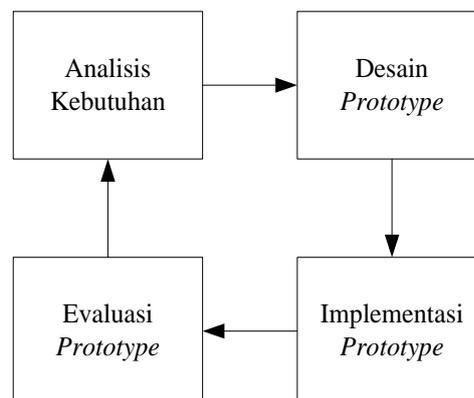
c. Implementasi *Prototype*

Setelah tahap desain aplikasi selesai, maka masuk ketahap berikutnya yaitu melakukan implementasi ke dalam bahasa pemrograman sesuai dengan desain aplikasi yang sebelumnya sudah dibuat.

d. *Construction*

Setelah program selesai, maka tahap selanjutnya adalah evaluasi dan melakukan pengujian terhadap *prototype* aplikasi yang telah dibuat. Pada tahap ini dilihat apakah masih ada kekurangan atau *error* pada *prototype*, apabila terdapat *error* atau kekurangan maka akan dicatat kemudian kembali lagi ke tahap nomor 1 untuk melakukan perbaikan. Proses berakhir ketika pada tahap evaluasi hingga tidak terdapat lagi *error* atau kekurangan.

Penggambaran model *prototype* dapat dilihat pada gambar Gambar 1.2.



Gambar 1.2 Model Prototyping

1.5.5 Metode Pengujian

Dalam tahap ini akan dilakukan pengujian akurasi dari pengaplikasian metode *Least Square Support Vector Machine* pada aplikasi *Optical Character Recognition* pada sertifikat. Perhitungan akurasi dari *Optical Character Recognition* dilakukan dengan menggunakan (*true value* dan *reference value*).

1.5.6 Penarikan Kesimpulan

Berdasarkan pada hasil dari pengujian maka dapat ditarik kesimpulan mengenai hasil dari penelitian yang dilakukan yaitu OCR dengan menggunakan LS-SVM.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi yang akan dibuat terbagi kedalam 5 Bab beserta pokok materinya. Sebagai gambaran umum, sistematika penyusunan skripsi yang akan ditulis adalah sebagai berikut :

BAB 1 PENDAHULUAN

Pendahuluan menjelaskan mengenai Latar Belakang Masalah, Identifikasi Masalah, Maksud dan Tujuan Penelitian, Batasan Masalah, Metodologi Penelitian, dan Sistematika Penulisan.

BAB 2 LANDASAN TEORI

Didalam bab ini membahas landasan teori yang menunjang dalam penelitian dalam menyelesaikan masalah. Seperti *Optical Character Recognition, Preprocessing, Sertifikat, Least Square Support Vector Machine (LS-SVM)*.

BAB 3 ANALISIS DAN PERANCANGAN

Pada bab ini membahas tentang analisis masalah, analisis metode yang akan digunakan, analisis kebutuhan non-fungsional yang diantaranya ada analisis kebutuhan perangkat lunak, perangkat keras dan analisis *User*, serta perancangan sistem yang mencakup perancangan komponen aplikasi.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini membahas tentang implementasi, terdiri atas implementasi perangkat keras, implementasi perangkat lunak, implementasi aplikasi, dan pengujian, yang terdiri atas pengujian *black box* dan validasi metode LS-SVM serta kesimpulan hasil pengujian apakah sudah sesuai dari hasil analisis dan perancangan yang telah dibuat.

BAB 5 KESIMPULAN DAN SARAN

Berisi kesimpulan dari hasil penelitian *Optical Character Recognition* Menggunakan Algoritma *Least Square Support Vector Machine* Pada Sertifikat dan saran-saran untuk pengembangan penelitian yang akan datang.