

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Optical Character Recognition (OCR) merupakan sebuah sistem komputer yang dapat membaca citra huruf dan angka, baik cetakan (printer atau mesin ketik) maupun tulisan tangan. Sistem pengenalan huruf ini dapat meningkatkan kemampuan atau kecerdasan pada sistem komputer[1]. *Optical Character Recognition* (OCR) merupakan solusi yang efektif untuk proses konversi dokumen cetak ke dokumen digital.

Sertifikat merupakan tanda atau surat keterangan tertulis dari pihak yang berwenang yang dapat digunakan sebagai bukti kepemilikan atau suatu kejadian. Sertifikat juga merupakan sebuah surat sebagai tanda pengakuan bahwa seseorang telah menguasai kompetensi tertentu atau telah mengikuti acara tertentu. Pada saat ini sertifikat tidak hanya dibutuhkan dalam format cetak (*hardcopy*), namun dibutuhkan juga file (*softcopy*) dari sertifikat tersebut untuk menunjang beberapa hal seperti keamanan, kecepatan pengiriman, dan keutuhan informasi yang terdapat dalam sertifikat. Dalam sertifikat terdapat unsur atau bagian, bagian-bagian tersebut meliputi kop atau logo organisasi, nomor surat, nama peserta, waktu dan tempat,serta pengesahan [2]. Penelitian untuk mendeteksi teks pada sertifikat sebelumnya telah dilakukan oleh Reza Yogi Andriana dan Indra Riyanto yang menyimpulkan bahwa terdapat masalah pada bagian *preprocessing*. Penggunaan metode *Support Vector Machine* pada penelitian Reza Yogi Andriana memang memiliki akurasi yang lebih tinggi dari pada penelitian Indra Riyanto tetapi masih perlu memperbaiki pada tahap *preprocessing* yang tepat sehingga dapat mempengaruhi pada pengenalan huruf pada karakter yang kurang jelas.

Pengenalan pengolahan citra merupakan salah satu kemampuan yang dimiliki oleh komputer. Sebagian besar pengenalan huruf atau deteksi teks dilakukan pada huruf abjad berupa tulisan tangan biasa dengan teknik *Optical Character Recognition* (OCR) yang dapat mendeteksi dan mengenali teks pada citra digital

menjadi format teks [3]. Pada penelitian sebelumnya, OCR sudah digunakan untuk berbagai macam objek, seperti tulisan tangan yang memiliki akurasi sebesar 81.81% [4]. Karakter aksara Jawa dengan hasil akurasi sebesar 93.6% [5], dan masih banyak lagi.

Pada penelitian serupa telah dilakukan oleh Muhammad Fajry dan Galih Hermawan yang mempunyai akurasi 46,61% untuk kasus sensitif dan 54,30% untuk kasus tidak peka [6]. Sedangkan untuk pengkategorian dengan *Support Vector Machine* mencapai akurasi 95.23% sampai 100% [7]. Pada penelitian lain yang dilakukan oleh Reza Yogi Andriana yaitu OCR pada sertifikat memiliki akurasi sebesar 83,27952% [8]. Riset serupa juga telah dilakukan oleh Indra Rianto yang memiliki akurasi 67% untuk *Recall* [9]. Namun, masih terdapat kekurangan dalam hal segmentasi untuk kasus karakter yang kurang jelas.

Pada penelitian ini, metode *Maximally Stable Extremal Regions* (MSER) digunakan untuk tahap segmentasi dan *Support Vector Machine* untuk mengklasifikasi jenis kata dan kalimat pada citra sertifikat. Pada penelitian sebelumnya, MSER telah diaplikasikan pada berbagai macam citra dan menunjukkan efektivitasnya, seperti pada penelitian mengenai deteksi teks dari gambar natural [10], deteksi plat nomor kendaraan [11], deteksi teks pada KTP [12]. Pada kasus di atas bahwa metode segmentasi MSER mampu mempengaruhi pengenalan karakter terhadap klasifikasi teks yang memperoleh akurasi sebesar 98.85% [11] dan 84.5% [12].

Maka, penelitian ini akan dibuat sebuah *prototype* dari penerapan *Maximally Stable Extremal Regions* dan *Support Vector Machine* pada pengenalan citra sertifikat. Selanjutnya, akan dilakukan perhitungan akurasi dengan mengukur banyaknya kata yang dapat diklasifikasikan.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang sudah diuraikan sebelumnya maka identifikasi masalah pada penelitian ini adalah bagaimana cara untuk mengimplementasikan metode *Maximally Stable Extremal Regions* dan *Support Vector Machine* untuk mengenali huruf pada sertifikat.

1.3 Maksud dan Tujuan

Berdasarkan identifikasi masalah tersebut 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 segmentasi *Maximally Stable Extremal Regions* (MSER) dan metode klasifikasi *Support Vector Machine* dalam mengklasifikasikan sebuah huruf pada sertifikat.

1.3.2 Tujuan

Berdasarkan masalah yang telah diuraikan, maka dapat disusun tujuan dari penelitian ini adalah mengukur tingkat keakuratan *Support Vector Machine* (SVM) dan *Maximally Stable Extremal Regions* dalam pengenalan sebuah karakter pada sertifikat.

1.4 Batasan Masalah

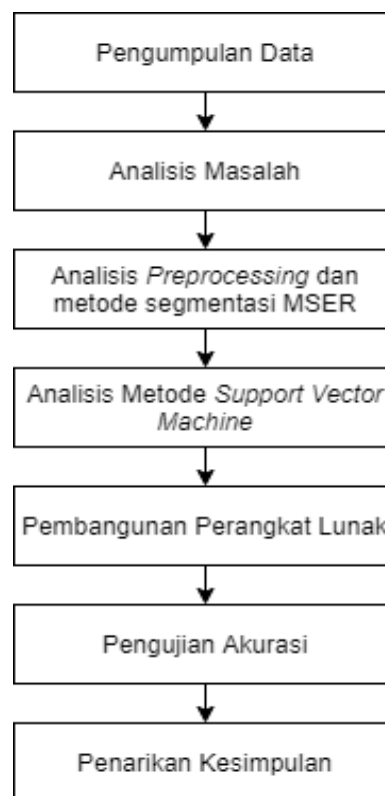
Pembatasan masalah diperlukan agar penelitian menjadi lebih terarah dan sesuai dengan tujuan yang sudah ditetapkan sebelumnya. Oleh karena itu beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dokumen yang di *scan* menggunakan format .jpg dan .png atau file e-sertifikat yang sudah diformat kedalam .jpg dan .png.
2. Sertifikat yang digunakan adalah sertifikat pada sebuah acara (seminar atau workshop).
3. Karakter yang digunakan bukan karakter huruf yang berjenis kursif.

4. Data latih didapatkan dengan cara menggunakan *word office* dengan setiap huruf mempunyai unsur A-Z,a-z,0-9 dengan jumlah 7.440 data latih.
5. Data uji didapatkan dari hasil *scanning* pada sertifikat atau e-sertifikat dengan jumlah 20 sertifikat.

1.5 Metode Penelitian

Metode penelitian yang digunakan pada penelitian ini adalah metode kuantitatif. Hasil dari penelitian ini adalah berupa persentase akurasi dengan mengakumulasikan tingkat kesesuaian antara input dengan hasil yang didapatkan. Berikut langkah-langkah yang dilakukan pada penelitian ini yang ditunjukkan pada Gambar 1.1 berikut:



Gambar 1.1 Metode Penelitian

Berikut ini adalah penjelasan dari alur metode penelitian pada Gambar 1.1 .

1. Pengumpulan Data

Adapun metode pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi Literatur dilakukan dengan membaca referensi dari jurnal, *paper*, dan buku terkait dengan sertifikat, deteksi teks, metode pengolahan citra, metode segmentasi citra dengan *Maximally Stable Extremal Regions* dan metode klasifikasi dengan *Support Vector Machine*.

2. Pengumpulan Dataset

Pengumpulan *dataset* dilakukan dengan mengambil data dari sampel pada citra sertifikat yang digunakan sebagai objek penelitian. Sampel tersebut nantinya akan digunakan sebagai data latih pada saat proses segmentasi citra dengan metode *Maximally Stable Extremal Regions* dan metode klasifikasi *Support Vector Machine*

2. Analisis Masalah

Analisis masalah yang dilakukan pada penelitian ini dibagi menjadi 2 yaitu:

- a. Analisis Data Masukan

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

- b. Analisis Kebutuhan Pembangunan

Analisis ini bertujuan untuk menganalisis kebutuhan seperti metode dalam *preprocessing*, *Maximally Stable Extremal Regions*, dan *Support Vector Machine*.

3. Analisis *Preprocessing* dan Metode *Maximally Stable Extremal Regions*

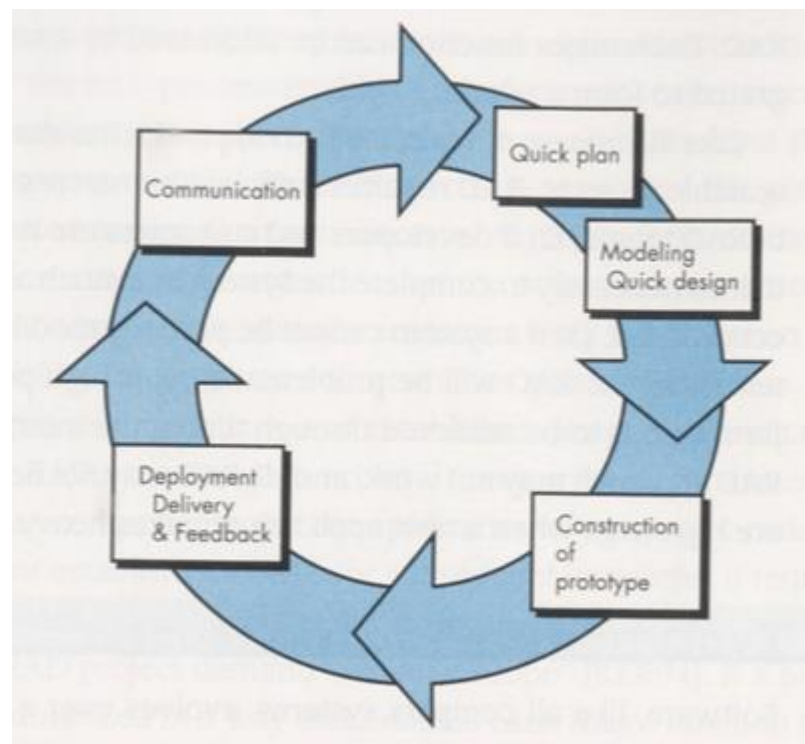
Pada tahap ini dilakukan analisis terhadap tahapan-tahapan pra-proses data masukan sebelum masuk ke dalam tahap klasifikasi. Adapun tahapan-tahapan *preprocessing* tersebut yaitu *grayscale*, *thresholding*, segmentasi, dan *resize*. kemudian pada tahap segmentasi akan dilakukan analisis metode segmentasi *Maximally Stable Extremal Regions*.

4. Analisis Metode *Support Vector Machine*

Tahapan ini dilakukan analisis terhadap metode klasifikasi *Support Vector Machine*. Metode SVM digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas.

1. Pembangunan Perangkat Lunak

Pada tahap ini dilakukan proses pembangunan perangkat lunak untuk mensimulasikan algoritma segmentasi dan klasifikasi yang digunakan pada penelitian ini. Metode pembangunan perangkat lunak yang digunakan pada penelitian adalah model *prototype*. Model *prototype* dipilih karena diketahui keberhasilan sistem dari evaluasi yang dilakukan. Jika *prototype* belum sesuai harapan, maka akan dilakukan perbaikan terhadap algoritma dan desain yang diimplementasikan pada sistem. Adapun gambaran metode pengembangan perangkat lunak dapat dilihat pada gambar 1.2



Gambar 1.2 Tahapan Metode Prototype

Berikut ini langkah-langkah pada proses Model *Prototype*:

1. *Communication*

Pada tahap ini dilakukan analisa fungsional dari kebutuhan perangkat lunak. Pada tahap ini pun dilakukan pengumpulan data-data yang dibutuhkan dalam penerapan metode *Maximally Stable Extremal Regions* (MSER) dan metode *Support Vector Machine* seperti membuat *dataset* sertifikat.

1. *Quick Plan*

Pada tahap ini dilakukan perencanaan awal yang akan dilakukan dalam membangun perangkat lunak pengenalan karakter pada sertifikat. Seperti menentukan waktu pengerjaan pembangunan perangkat lunak, membuat target pengerjaan dan menentukan jumlah dataset awal yang akan digunakan pada saat pelatihan dan pengujian.

2. *Modeling Quick Design*

Tahap ini dilakukan pemodelan yang akan dibuat pada pembangunan perangkat lunak, pemodelan tersebut antara lain terdiri dari pemodelan untuk *preprocessing*, pemodelan algoritma MSER dan *Support Vector Machine*, dan pemodelan perangkat lunak. Pemodelan perangkat lunak yang digunakan pada penelitian ini adalah UML (*Unified Model Language*). Selain itu pada tahap ini dilakukan juga tahap perancangan *mockup* antarmuka yang akan diimplementasikan pada tahap konstruksi

3. *Construction of Prototype*

Tahap ini dilakukan pengkodean dari hasil pemodelan yang sudah dilakukan sebelumnya. Pada penelitian ini pengkodean dilakukan dengan menggunakan bahasa pemrograman *Python*.

4. *Deployment Delivery & Feedback*

Tahap ini dilakukan pengujian terhadap perangkat lunak dan performansi metode MSER dan *Support Vector Machine* serta evaluasi terhadap langkah-langkah yang telah dilakukan dalam penerapan MSER dan *Support Vector Machine* serta perangkat lunak yang telah dibangun.

1. Pengujian Akurasi

Tahap ini merupakan tahap pengujian dari metode *Maximally Stable Extremal Regions* dan metode *Support Vector Machine*. Hasil dari pengujian ini berupa persentase akurasi dari metode *Support Vector Machine*.

1. Penarikan Kesimpulan

Pada tahap ini dilakukan pengambilan kesimpulan dari seluruh tahapan-tahapan yang dilakukan pada penelitian ini.

1.6 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mulai dari permasalahan hingga solusi atau pemecahannya. Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab 1 ini berisi tentang latar belakang masalah, identifikasi masalah, perumusan masalah, maksud dan tujuan, batasan masalah, dan metodologi penelitian, serta sistematika penulisan sebagai gambaran dalam penelitian ini.

BAB 2 LANDASAN TEORI

Bab 2 ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan topik penelitian yang akan dilakukan dan hal-hal yang berguna dalam proses analisis permasalahan serta tinjauan terhadap penelitian-penelitian serupa yang pernah dilakukan sebelumnya.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab 3 ini berisi tentang analisa dan perancangan dari sistem yang akan dibangun, yang meliputi analisis sistem, analisis masukan, analisis *preprocessing*, analisis klasifikasi, analisis keluaran, analisis kebutuhan non fungsional, dan perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab 4 ini memaparkan tentang hasil implementasi dari analisa dan perancangan yang telah dilakukan sebelumnya serta hasil pengujian dari sistem yang dibangun. Pengujian sistem dilakukan dengan menggunakan pengujian *black box*.

BAB 5 KESIMPULAN DAN SARAN

Bab 5 ini menjelaskan tentang kesimpulan dari hasil penelitian yang telah dilakukan dan saran untuk penelitian lebih lanjut.



