

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Optical character recognition (OCR) adalah sebuah sistem komputer yang dapat membaca huruf, baik yang berasal dari sebuah pencetak (printer atau mesin ketik) maupun yang berasal dari tulisan tangan. OCR adalah aplikasi yang menerjemahkan gambar karakter (image character) menjadi bentuk teks dengan cara menyesuaikan pola karakter per baris dengan pola yang telah tersimpan dalam database aplikasi. Hasil dari proses OCR adalah berupa teks sesuai dengan gambar output scanner dimana tingkat keakuratan penerjemahan karakter tergantung dari tingkat kejelasan gambar dan metode yang digunakan [1].

Penelitian sebelumnya mengenai Optical Character Recognition telah dilakukan sebelumnya oleh peneliti sebelumnya dengan menggunakan berbagai metode, pada penelitian Aldi Setiawan, Herry Sujaini, Arif Bijaksana PN (2017), implementasi Optical Character Recognition berbasis android, penelitian ini mengangkat masalah utama yaitu tidak efisiennya penggunaan kamus cetak untuk mencari arti dari satu kata demi satu kata, dan untuk penerjemah elektronik akan menerjemahkan kata atau kalimat yang diketik oleh user melalui keyboard atau keypad sebagai inputan dan diperoleh hasil pengujian recall dan precision aplikasi, diperoleh bahwa aplikasi cukup baik dalam menangkap kata masukan dengan beberapa font dan ukuran font antara 12 sampai 16. [2]

Pada penelitian yang termasuk ke dalam text processing menggunakan RSVM Epa Suryanto dan Santi Wulan Purnami (2015) Metode SSVM membutuhkan waktu running yang juga hampir sama dengan RSVM, tetapi untuk jumlah data lebih dari 1000 waktu yang dibutuhkan metode SSVM mengalami peningkatan yang tinggi. Pada metode RSVM waktu yang dibutuhkan dengan berbagai jumlah data lebih cepat dibandingkan SSVM. Performansi RSVM lebih baik daripada SSVM untuk jumlah data yang relatif besar (lebih dari 1000). Sedangkan pada data yang relatif kecil (kurang dari 1000) kedua metode memberikan performansi yang sama. Akurasi RSVM pada data simulasi lebih besar dari 99%. [3]

Penelitian Santi Wulan Purnami, Jasni Mohamad Zain, Tutut Heriawan (2011) Memberikan alternatif pada metode SVM yang tidak efisien untuk memecahkan masalah optimasi pada data besar yaitu dengan memakai metode RSVM sebagai modifikasi dari metode SVM. Dari hasil perbandingan antara metode RSVM dan SSVM, metode RSVM mampu untuk mengklarifikasi jumlah data diatas 8000 data, sedangkan pada metode SSVM fungsi kernel mengalami out of memory bahkan sebelum memulai pencarian solusi. [4]

Penelitian Kuang Ming Lin dan Chih Jen Lin (2003) membahas tentang perbandingan metode RSVM dan SVM yang merupakan modifikasi dari metode SVM. Hasil penelitian menjelaskan bahwa hasil akurasi kecepatan RSVM lebih cepat daripada SVM. Dan metode RSVM dapat digunakan pada proses klarifikasi pada data besar sedangkan SVM tidak bisa [5]

Penelitian Vincent Limountha membahas tentang esktraksi informasi pada dokumen surat keputusan. Kasus yang diangkat dari penelitian ini adalah untuk mengetahui tingkat akurasi dari metode *Conditional Random Fields* (CRF). Hasil dari penelitian menjelaskan bahwa hasil akurasi kecepatan CRF sebesar 91,73 %, tingkat akurasi ini diperoleh oleh nilai parameter awal pada proses pelatihan. [6]

Berdasarkan penelitian sebelumnya penerapan RSVM menghasilkan nilai akurasi yang baik, maka kali ini perlu dilakukan sebuah penelitian untuk membangun sebuah sistem ekstraksi informasi pada surat keputusan menggunakan algoritma RSVM dengan Metode OCR. Dengan penelitian ini diharapkan dapat menemukan hasil akhir apakah penggunaan OCR dengan metode algoritma RSVM memiliki nilai akurasi yang baik.

1.2 Identifikasi Masalah

Berdasarkan latar belakang masalah yang diuraikan diatas, masalah yang dapat teridentifikasi yaitu dibutuhkannya hasil akurasi dari sistem ekstraksi informasi dengan menggunakan Algoritma RSVM untuk dokumen surat keputusan.

1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk Mengimplementasikan Optical Character Recognition pada dokumen Surat Keputusan.

Berdasarkan rumusan masalah yang telah dipaparkan diatas, maka tujuan dilakukannya penelitian ini untuk mengetahui nilai akurasi, presisi, dan *recall* algoritma RSVM dan Metode OCR dalam implementasi ekstraksi informasi pada Surat Keputusan.

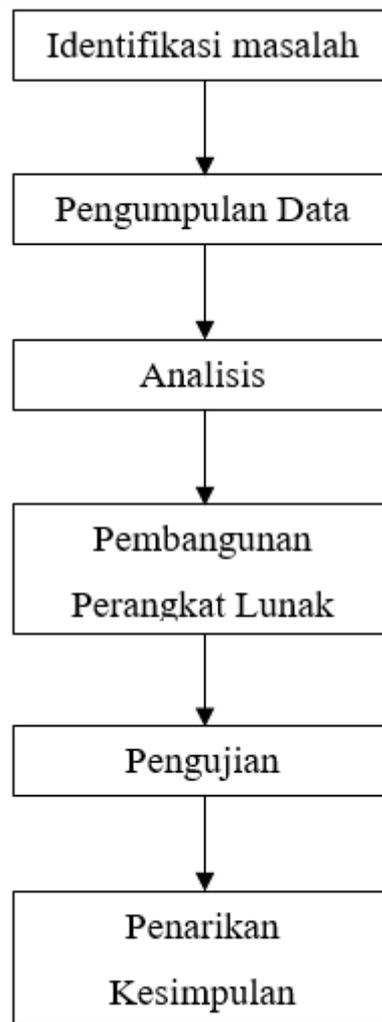
1.4 Batasan Masalah

Batasan masalah pada penelitian yang akan dilakukan adalah:

1. Penelitian ini hanya melakukan pengambilan data berupa angka dan huruf pada Surat Keputusan.
2. Data yang di gunakan dalam penelitian ini hanya badan surat, Kop Surat dan tanda tangan tidak termasuk dalam proses ekstraksi.
3. Data masukan yang digunakan pada penelitian ini adalah Surat Keputusan yang ditujukan untuk satu orang.
4. Jenis dokumen yang digunakan untuk data training pada penelitian ini berformat *.jpg.

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. Identifikasi Masalah

Berdasarkan fenomena tersebut maka rumusan masalah adalah karakteristik pada dokumen surat keputusan tiap instansi berbeda sehingga dibutuhkan klasifikasi untuk mengekstraksi informasi didalamnya. Metode klasifikasi yang akan digunakan adalah RSVM dikarenakan iya memiliki hasil yang baik dalam ekstraksi informasi.

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 :

a. Studi Pustaka

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

b. Dataset

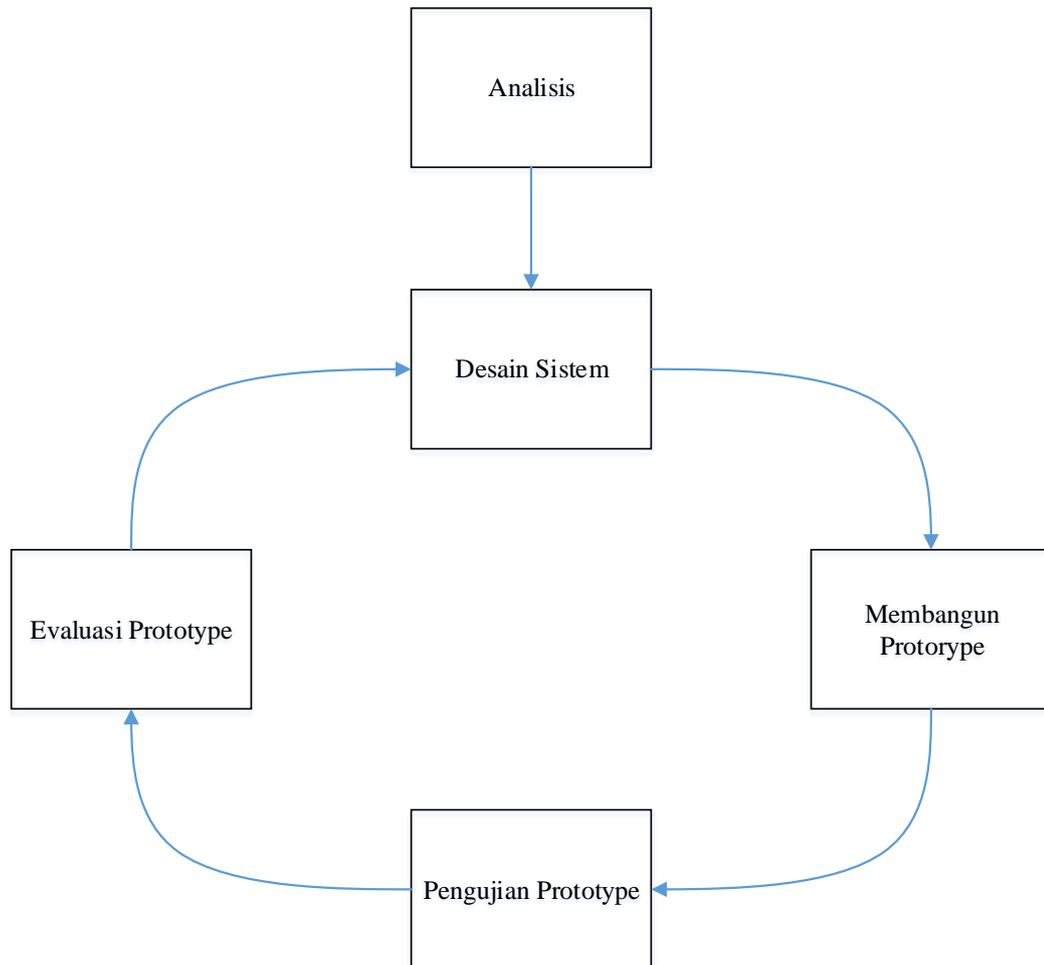
Pengumpulan data dengan mengumpulkan bahan dari internet, literatur, maupun sumber-sumber yang memiliki data yang akan menjadi bahan dalam penelitian.

3. Analisis

Tahap analisis yang dilakukan adalah analisis data masukan yang akan digunakan, analisis tahap *preprocessing*, sampai analisis ekstraksi dengan menggunakan metode OCR dan algoritma RSVM.

4. Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak pada penelitian ini dengan metode *prototype*.



Gambar 1.2 Metode Pembangunan Perangkat Lunak

Pemodelan sistem yang digunakan menggunakan pemodelan *Prototype*. Berikut tahapan pada model *Prototype* .

a. Analisis Sistem

Mengumpulkan kebutuhan-kebutuhan untuk sistem yang dibuat kemudian dianalisis sesuai dengan kebutuhan penelitian seperti dokumen pendukung dan informasi tambahan yang dibutuhkan.

b. Desain

Menjabarkan informasi yang didapat dan telah dianalisis secara rinci seperti kebutuhan *software* dan *hardware*

c. Pembangunan *Prototype*

Tahap penerjemahan desain ke dalam bahasa pemrograman yang berisikan pemecahan masalah pada penelitian.

d. Pengujian

Menguji apakah sistem yang dibangun dapat memecahkan masalah atau tidak, sesuai dengan apa yang telah dirancang sebelumnya.

e. Evaluasi *Prototype*

Pada tahap ini sistem akan mengalami perubahan-perubahan tergantung permintaan atau kebutuhan pengguna.

5. Pengujian

Pengujian yang dilakukan pada penelitian ini menggunakan *confussion matrix* dengan menghitung *precision*, *recall*, dan akurasi.

6. Penarikan Kesimpulan

Penarikan kesimpulan dilakukan setelah pengujian dilakukan. Hasil dari pengujian dijadikan bahan untuk menarik kesimpulan, apakah RSVM mempunyai nilai akurasi yang bagus atau tidak dalam implementasi ekstraksi informasi pada Surat Keputusan.

1.6 Sistematika Penulisan

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

BAB I 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 II LANDASAN TEORI

Pada bab ini akan menjelaskan mengenai objek dari penelitian, dan teori – teori diantaranya yaitu teori RSVM, teori OCR, teori Pengolahan Citra, serta teori pendukung yang berhubungan dengan penelitian ini serta perangkat lunak pendukung yang digunakan dalam penelitian ini.

BAB III. ANALISIS DAN PERANCANGAN SISTEM

Pada bab ini membahas tentang analisis masalah dan kebutuhan mengenai metode *Optical Character Recognition* dan algoritma RSVM dalam men ekstrak

dokumen. Penjelasan proses dalam ekstraksi informasi secara matematis. Pada bab ini juga dibahas tentang perancangan UML dan perancangan antarmuka.

BAB IV. IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab ini berisi mengenai hasil dari implementasi dari analisis yang telah dilakukan di BAB 3 dan pengujian akurasi metode OCR dan algoritma RSVM yang digunakan untuk ekstraksi informasi pada surat keputusan.

BAB V. KESIMPULAN DAN SARAN

Berisi kesimpulan dari hasil penelitian Ekstraksi Informasi pada Surat Keputusan Menggunakan metode *Optical Character Recognition* dan algoritma RSVM dan saran-saran untuk pengembangan penelitian yang akan datang.