

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

OCR atau *Optical Character Recognition* adalah salah satu pengaplikasian teknologi yang digunakan untuk pengenalan karakter atau pola[1]. 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 hasilnya bisa menjadi sebuah teks yang utuh atau biasa disebut *Text Recognition*. OCR sendiri bekerja dengan melakukan pelatihan terhadap karakter yang akan dikenali menggunakan algoritma tertentu yang memiliki beragam tingkat akurasi tentunya[2].

Sebelumnya Arbi Insani melakukan sebuah penelitian tentang ekstraksi informasi pada dokumen teks novel dengan memanfaatkan OCR. Adapun jenis informasi yang diekstrak meliputi judul novel, nama penulis, tahun terbit, penerbit, dan ISBN. Namun didalam penelitiannya Arbi memiliki kekurangan terkait pengenalan pada judul novel, dimana hasil yang didapat tidak sesuai dengan judul novel yang sesungguhnya. Dibuktikan dengan hasil yang diperoleh dimana 23 dari 50 judul novel tidak dapat dikenali dengan benar [3]. Hal ini disebabkan karena kebanyakan huruf pada judul novel berjenis *cursive*. Huruf *cursive* sendiri merupakan salah satu jenis huruf unik karena bentuknya mirip dengan tulisan tangan yang memiliki kesan personal dan berbeda dibanding jenis huruf lainnya[4].

Solusi yang bisa ditawarkan untuk menyelesaikan permasalahan tersebut adalah dengan memanfaatkan metode – metode yang ada seperti *Artificial Neural Network*, *Backpropagation*, *KNN*, dan *Support Vector Machine* yang bisa digunakan untuk membangun aplikasi pengenalan citra pada judul novel berjenis

cursive. Dari metode yang sudah disebutkan tadi, *Support Vector Machine* (SVM) merupakan metode yang paling sering digunakan karena memiliki performansi rata-rata terbilang lebih baik diberbagai bidang aplikasi seperti *bioinformatics*, klasifikasi teks dan khususnya pada pengenalan tulisan atau huruf[5].

Penggunaan algoritma SVM sendiri sebelumnya sudah beberapa kali digunakan pada pembangunan aplikasi OCR. Salah satu contohnya adalah penelitian yang dilakukan oleh Puneet Jhaji dimana SVM digunakan untuk pengenalan huruf india menggunakan kernel RBF yang dikombinasikan dengan ekstraksi fitur *zoning* yang memiliki akurasi 73.83% [6]. Selain itu juga melakukan penelitian terkait SVM yang digunakan untuk pengenalan tulisan jawi menggunakan kernel *quadratic* dengan ekstraksi fitur *NRC* sebesar 74.47% [7].

Jenis kernel yang digunakan dan jenis ekstraksi fitur akan berpengaruh seberapa besar pada tingkat performansi dari SVM itu sendiri. Kombinasi dari penggunaan jenis kernel dan jenis ekstraksi fitur yang dirasa paling cocok adalah kernel RBF dengan fitur ekstraksi diagonal. Dibuktikan dengan beberapa penelitian sebelumnya, salah satunya dilakukan oleh Chandra Prakash yang mampu menghasilkan akurasi sebesar 93.06% [8] untuk mengenali karakter Gurmukhi (tulisan India). Selain itu juga pernah dilakukan penelitian oleh Rani Rajneesh untuk kasus yang sama dan menghasilkan akurasi sebesar 95.34% [6].

Berdasarkan pemaparan yang sudah disampaikan, maka pada penelitian ini ditujukan untuk menerapkan metode SVM menggunakan kernel RBF yang dikombinasikan dengan fitur ekstraksi diagonal untuk pengenalan teks pada judul novel berjenis *cursive*.

1.2 Perumusan Masalah

Berdasarkan uraian yang sudah disampaikan pada latar belakang sebelumnya, didapatkan perumusan masalah yaitu seberapa besar tingkat akurasi dari algoritma SVM menggunakan kernel RBF yang dikombinasikan dengan fitur ekstraksi diagonal terhadap pengenalan teks pada judul novel berjenis *cursive*.

1.3 Maksud dan Tujuan

Adapun maksud dan tujuan dari penelitian ini adalah sebagai berikut.

1.3.1 Maksud

Maksud dari penelitian ini adalah untuk membuat aplikasi yang dapat mengenali teks pada judul novel berjenis *cursive* menggunakan algoritma SVM yang dikombinasikan dengan fitur ekstraksi diagonal.

1.3.2 Tujuan

Tujuan dari penelitian ini adalah mengetahui tingkat keakuratan algoritma SVM yang dikombinasikan dengan fitur ekstraksi diagonal untuk pengenalan judul novel berjenis *cursive*.

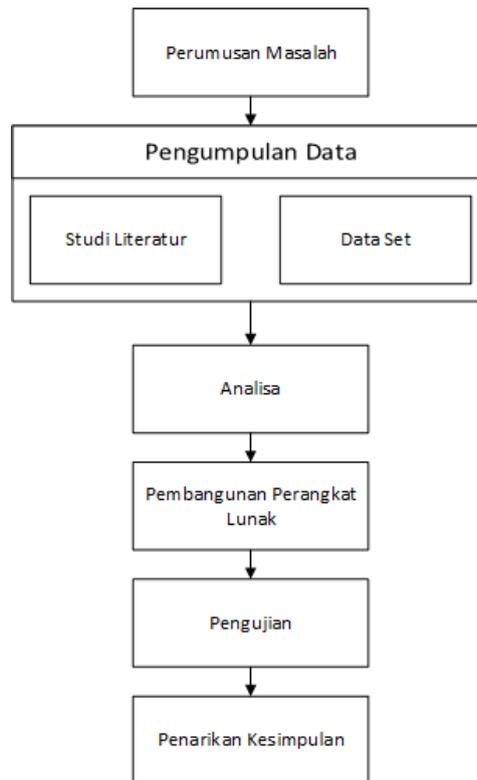
1.4 Batasan Masalah

Agar masalah yang sedang ditinjau lebih terarah dan mencapai tujuan yang telah ditentukan, maka akan dibatasi masalah sebagai berikut yang terbagi kedalam beberapa aspek diantaranya.

1. Data yang digunakan dalam pengolahan adalah citra digital dari novel hasil scan berformat .jpg.
2. Bagian novel yang akan dikenali hanya pada bagian judulnya saja.
3. Jenis ekstraksi fitur citra yang digunakan adalah *Diagonal Based Features*.
4. Keluaran sistem berupa teks digital.

1.5 Metode Penelitian

Dalam penelitian ini, metode yang digunakan adalah metode deskriptif. Metode ini digunakan karena pembangunan sistem yang dilakukan berdasarkan dari gambaran fenomena-fenomena yang ada, serta sistematis dan terstruktur. Berikut adalah langkah - langkah yang dilakukan dalam penelitian.



Gambar 1. 1 Alur Penelitian

Berdasarkan alur penelitian pada gambar 1.1 dapat di jelaskan sebagai berikut:

1.5.1 Perumusan Masalah

Tahap pertama dalam penelitian adalah perumusan masalah. Berkaitan dengan kegiatan menganalisa permasalahan yang muncul mengenai kendala pada proses pengenalan huruf atau karakter pada judul novel yang berjenis *cursive*. Serta apa saja faktor – faktor yang mempengaruhi terhadap masalah yang ada.

1.5.2 Pengumpulan Data

Tahapan kedua adalah pengumpulan data. Metode pengumpulan data yang dilakukan pada penelitian ini yaitu :

a. Studi Literatur

Studi Literatur dilakukan dengan membaca jurnal, paper, buku dan sumber lain terkait dengan topik OCR, metode SVM, fitur ekstraksi

diagonal dan hal lain yang berkaitan dengan topik penelitian. Dengan tujuan untuk pemahaman dasar peneliti terhadap kasus yang diteliti.

b. Pengumpulan Data Set

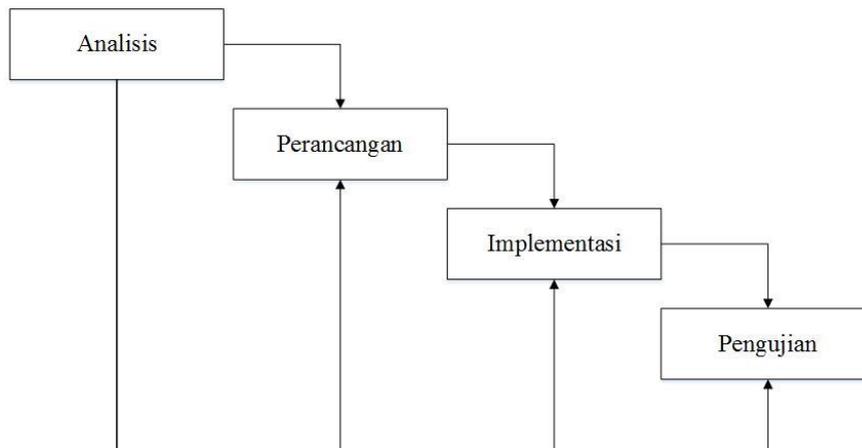
Pada tahap ini diambil pengumpulan data novel dengan 50 sample novel dan beragam jenis *cursive font*. Data sampel ini digunakan sebagai data uji, sedangkan data latih diambil dari 30 jenis huruf yang mirip pada data uji mulai dari huruf a-z dan A-Z.

1.5.3 Analisa

Selanjutnya setelah data terkumpul kemudian masuk ke dalam tahap analisa, dimana dalam tahapan ini data literatur dan referensi yang sudah terkumpul dipelajari terkait metode yang akan digunakan yaitu SVM, ekstraksi fitur diagonal bagaimana alur proses serta bagaimana cara kerjanya, fitur apa saja yang dibutuhkan dan hal lainnya terkait pembangunan sistem termasuk menganalisa bagaimana nantinya sistem yang dibuat akan berjalan mulai dari tahapan *input* masuk ke tahap proses hingga hasil berupa *output* dari sistem berupa teks hasil pengenalan. Kemudian data set yang terkumpul dibagi atau dipisahkan mana yang akan dijadikan data latih dan mana yang akan dijadikan data uji.

1.5.4 Pembangunan Perangkat Lunak

Dalam pengembangan sistem, metode yang digunakan adalah metode *waterfall* karena proses dalam membangun sistem yang dilakukan secara terurut. Tahap perancangan tidak dapat dilakukan jika tahap analisis belum dilakukan. Begitupun dengan tahap implementasi dan pengujian tidak dapat dilakukan jika tahap analisis dan perancangan belum dilakukan. Adapun tahapannya dapat dilihat pada gambar 1.2.



Gambar 1.2 Diagram Waterfall Sommerville

a. Analisis Kebutuhan Perangkat Lunak

Langkah ini merupakan tahap analisis terhadap kebutuhan perangkat lunak, dan tahap dalam pengumpulan kebutuhan penelitian secara menyeluruh yaitu pengenalan teks karakter pada judul novel. Bagaimana nantinya sistem akan dibangun dan bekerja dengan metode serta algoritma yang diterapkan. Data yang diambil adalah data novel dalam format gambar atau citra, dan mengumpulkan data – data tambahan baik yang ada di jurnal, artikel, maupun dari sumber lainnya.

b. Perancangan

Setelah kebutuhan data dan pemrosesan telah terkumpul, maka tahap selanjutnya adalah perancangan arsitektur sistem yang akan dibuat. Mulai dari perancangan struktur tabel, perancangan struktur menu dan perancangan tampilan antarmuka.

c. Implementasi

Setelah perancangan sistem secara keseluruhan, maka tahap selanjutnya adalah mengimplementasikan kedalam bahasa pemrograman algoritma yang digunakan. Tahapan dimana algoritma *Support Vector Machine* (SVM) digunakan dalam menyelesaikan permasalahan mengenai pengenalan karakter tulisan pada judul novel.

d. Pengujian sistem

Tahapan ini digunakan untuk mengintegrasikan program dan pengujian tujuan, apabila program yang dibuat berjalan sesuai dengan harapan, dan sesuai dengan pendekatan aturan maka pengenalan karakter tulisan judul novel siap digunakan oleh pengguna sistem.

1.5.5 Pengujian

Tahap selanjutnya adalah pengujian terkait nilai performansi dari algoritma SVM yang dikombinasikan dengan ekstraksi fitur diagonal untuk pengenalan teks pada judul novel.

1.5.6 Penarikan Kesimpulan

Ini merupakan tahap akhir dalam penelitian dengan menyimpulkan hasil keseluruhan penelitian yang sudah dilakukan baik terkait sistem yang dibuat, penggunaan metode yang digunakan terhadap kasus hingga saran untuk penelitian kedepannya.

1.6 Sistematika Penulisan

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

BAB 1 PENDAHULUAN

Bab ini merupakan bab yang membahas latar belakang masalah, perumusan masalah, maksud dan tujuan, batasan masalah, metodologi penelitian, serta sistematika penulisan tentang proses pengenalan karakter.

BAB 2 LANDASAN TEORI

Bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan novel, pengenalan karakter, kecerdasan buatan, bahasa pemrograman, kajian pendekatan SVM, pendekatan Diagonal Feature Extraction, uraian mengenai proses preprocessing, uraian singkat UML, flowchart, jaringan semantic dan pengujian.

BAB 3 ANALISIS DAN PERANCANGAN

Bab ini menjelaskan analisis dan perancangan sistem yang dibangun. Analisis sistem terdiri dari analisis masalah, analisis data, analisis metode, sistem yang sedang berjalan, analisis kebutuhan non fungsional yang terdiri dari analisis kebutuhan perangkat keras, analisis kebutuhan perangkat lunak.

BAB 4 IMPLEMENTASI DAN PENGUJIAN PERANGKAT LUNAK

Bab ini menjelaskan implementasi dari perangkat lunak yang dibangun. Implementasi perangkat lunak dilakukan berdasarkan kebutuhan analisis dan perancangan perangkat lunak yang sudah dilakukan. Dari hasil implementasi kemudian dilakukan pengujian berdasarkan pendekatan penggunaan algoritma SVM yang dikombinasikan dengan *diagonal features extraction* pada analisis kebutuhan perangkat lunak yang menjelaskan apakah sudah benar-benar sesuai dengan analisis dan perancangan yang telah dilakukan

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan hasil penelitian yang telah dilakukan dan saran yang diharapkan dapat dikembangkan untuk penelitian selanjutnya.