

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Optical Character Recognition (OCR) adalah bagian dari perangkat lunak yang mengubah teks cetakan dan gambar menjadi bentuk digital sehingga dapat dimanipulasi oleh mesin. Tidak seperti otak manusia yang memiliki kemampuan dengan mudah mengenali teks / karakter dari suatu gambar, mesin tidak cukup cerdas untuk melihat informasi tersedia dalam gambar[1]. Oleh karena itu, sejumlah besar upaya penelitian telah diajukan untuk mencoba mengubah gambar dokumen menjadi format yang dapat dimengerti oleh mesin.

Dalam penelitian Reza Yogi Andria yang berjudul *Optical Character Recognition* (OCR) Menggunakan *Support Vector Machine* (SVM) Dan *Zoning* Pada Sertifikat diperoleh akurasi terbaik sebesar 83,27952%[2]. Kemudian pada penelitian Aditya Surya Wijaya, Nurul Chamidah dan Mayanda Mega Santoni yang berjudul Pengenalan Karakter Tulisan Tangan Dengan *K-Support Vector Nearest Neighbor* mendapatkan bahwa metode K-SVNN dapat meningkatkan akurasi dibandingkan dengan tanpa menggunakan K-SVNN pada pengenalan karakter angka dimana akurasi mencapai 85.81%[3].

Sertifikat merupakan suatu bentuk desain dan dicetak yang didalamnya memiliki struktur dan keterangan informasi tertentu dari suatu instansi atau seseorang yang berwenang[4]. Sertifikat memiliki informasi seperti nama penerima, nama pengirim, tanggal dan sebagainya. Untuk mendapatkan informasi yang terkandung dalam sertifikat, maka muncul kebutuhan untuk mengekstraksi informasi dari data sertifikat yang berupa gambar menjadi sebuah teks digital sehingga scan sertifikat akan dijadikan objek dalam penelitian ini.

K-Nearest Neighbor (K-NN) merupakan metode klasifikasi yang populer dalam data mining dan sangat kompetitif dalam klasifikasi banyak kelas[5], Namun, KNN lebih banyak menggunakan memori pada proses pelatihan dan kecepatan klasifikasi rendah[6]. Oleh karena itu, diperlukan modifikasi pada

metode KNN untuk mengatasi masalah memori dan waktu. Pada penelitian ini digunakan suatu metode untuk mereduksi data latih, sehingga diharapkan penggunaan memori dapat berkurang dan proses klasifikasi pengujian cepat tanpa mengabaikan hasil akurasi[3]. Reduksi *K-Support Vector Nearest Neighbor* (K-SVNN) merupakan metode hasil evolusi dari K-NN yang efektif untuk mengurangi memory dan waktu komputasi [7]. K-SVNN akan mereduksi data latih terlebih dahulu yang diharapkan dapat mengurangi waktu komputasi untuk selanjutnya dilakukan klasifikasi[8].

Pada penelitian sebelumnya K-SVNN mempunyai akurasi yang lebih baik daripada *Decision Tree* dan *Naïve Bayes* dengan selisih lebih baik paling kecil 18.20%[7]. Kemudian K-SVNN mempunyai akurasi yang relatif lebih baik dan waktu pelatihan yang relatif lebih singkat daripada ANN-EBP dan SVM[9]. Selain itu penggunaan K-SVNN juga menunjukkan bahwa waktu pelatihan berkurang sebesar 15% hingga 80%, sedangkan perbedaan akurasi prediksi ada penurunan sebesar 0% hingga 4.76%[10]. Maka pada penelitian ini digunakan klasifikasi K-NN dan Reduksi K-SVNN sebagai metode untuk pengenalan karakter pada dokumen sertifikat.

1.2. Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, maka identifikasi masalah dalam penelitian ini adalah sulitnya mengenali setiap huruf dan angka dengan ukuran dan jenis font yang berbeda-beda pada setiap dokumen sertifikat.

1.3. Maksud dan Tujuan

Berdasarkan permasalahan tersebut, maka maksud dari penelitian ini adalah untuk mengimplementasikan metode klasifikasi K-NN dan reduksi K-SVNN pada kasus ekstraksi informasi dokumen sertifikat. Adapun tujuannya yaitu untuk mengukur tingkat akurasi metode K-NN dan K-SVNN dalam mengenali setiap huruf dan angka pada dokumen sertifikat.

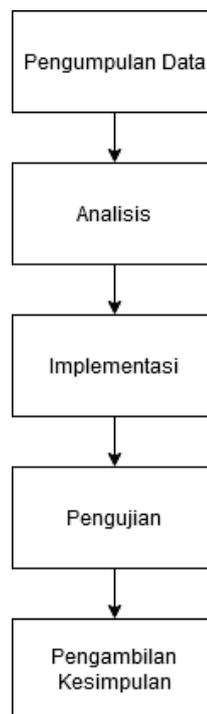
1.4. Batasan Masalah

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data masukan diperoleh dari umum dan internet.
2. Data masukan yang digunakan menggunakan format .jpg dan .png
3. Dokumen sertifikat harus memiliki kualitas gambar yang baik agar dapat terdeteksi pada praproses dan segmentasi.
4. Data yang dikumpulkan berupa dataset huruf (A - Z), (a - z), dan angka (0 - 9).
5. Data yang dilakukan pengenalan hanya huruf dan angka, untuk simbol dan *noise* tidak dilakukan proses pengenalan.

1.5. Metode Penelitian

Metode penelitian yang dipakai dalam penelitian ini memiliki lima tahap yaitu tahap pengumpulan data, analisis, implementasi, pengujian, dan pengambilan kesimpulan yang dapat dilihat pada gambar 1.1 berikut ini:



Gambar 1.1 Alur Penelitian

1.5.1. Pengumpulan Data

Metode pengumpulan data dalam penelitian ini adalah studi literatur yang dilakukan dengan cara mengumpulkan data dan informasi dari literatur, jurnal,

paper, buku, serta bacaan lainnya yang berkaitan dengan OCR, metode klasifikasi K-NN dan reduksi K-SVNN yang akan digunakan dalam penelitian.

1.5.2. Analisis

Analisis yang dilakukan pada penelitian ini dibagi menjadi dua yaitu:

1. Analisis Data Masukan

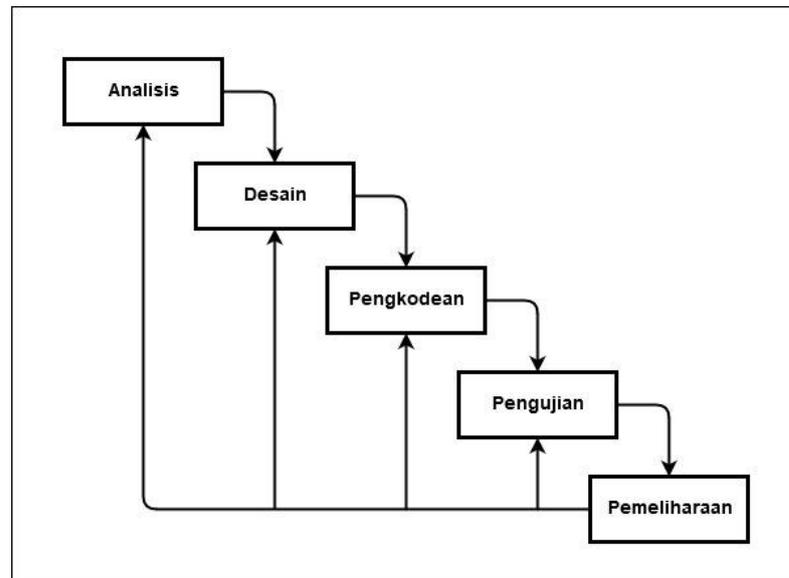
Analisis ini bertujuan untuk menganalisis data masukan yang akan diproses, yaitu berupa karakter-karakter pada sebuah jenis huruf tertentu.

2. Analisis Kebutuhan Pembangunan

Analisis ini bertujuan untuk menganalisis kebutuhan seperti metode dalam preprocessing, metode K-NN dan reduksi K-SVNN.

1.5.3. Implementasi

Implementasi dalam penelitian ini menggunakan model *Waterfall*. *Waterfall* adalah model klasik yang bersifat sistematis, berurutan dalam membangun software[11]. Pada model pengembangan sistem metode *waterfall*, sebuah pengembangan sistem dilakukan berdasarkan urutan analisis, desain, pengkodean, pengujian, dan berakhir pada tahap pemeliharaan. Disebut sebagai metode *waterfall* dikarenakan tahapan dan juga urutan dari metode yang dilakukan merupakan jenis metode yang berurutan dan berkelanjutan, seperti yang terlihat pada gambar 1.1 berikut ini:



Gambar 1.2 Metode *Waterfall*

1. Analisis

Langkah ini merupakan analisis terhadap kebutuhan perangkat lunak, dan tahap dalam pengumpulan kebutuhan penelitian secara lengkap tentang ekstraksi informasi pada sertifikat. Data yang diambil adalah data sertifikat dalam format .jpg atau .png dan mengumpulkan data-data tambahan baik yang ada di jurnal, artikel, maupun dari internet.

2. Desain

Setelah kebutuhan data dan pemroses telah dikumpulkan, maka tahap selanjutnya adalah perancangan arsitektur sistem secara keseluruhan. Untuk kebutuhan perangkat lunak dalam pengolahan data diperlukan *software* diantaranya penyunting kode sumber.

3. Pengkodean

Setelah perancangan sistem secara keseluruhan, maka tahap selanjutnya adalah mengimplementasikan kedalam bahasa pemrograman. Tahap ini digunakan untuk merancang program yang telah didesain sebelumnya.

4. Pengujian

Tahapan ini digunakan untuk mengintegrasikan program dan pengujian tujuan, apabila program yang dibuat berjalansesuai dengan harapan, dan sesuai dengan pendekatan aturan maka ekstraksi informasi sertifikat siap digunakan oleh pengguna sistem.

5. Pemeliharaan

Pada tahapan ini sistem sudah siap untuk digunakan, lalu dilakukan pengembangan sistem seperti penambahan fitur dan fungsi baru yang mungkin akan muncul untuk kebutuhan.

1.5.4. Pengujian

Pada tahapan ini adalah menghitung keakuratan dari klasifikasi K-NN dan reduksi K-SVNN dalam mengenali karakter pada dokumen citra sertifikat dokumen sertifikat. Metode yang digunakan untuk melakukan pengujian sistem adalah menggunakan metode *Black Box*.

1.5.5. Penarikan Kesimpulan

Langkah kelima adalah penarikan kesimpulan berdasarkan pengujian yang telah dilakukan sebelumnya. Penarikan kesimpulan dari hasil penerapan metode K-NN da reduksi K-SVNN.

1.6. Sistematika Penulisan

Sistematika Penulisan penelitian ini dibagi dalam beberapa bab dengan pokok-pokok permasalahannya. Sistematika penulisan secara umum dari laporan ini sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian serta sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan penelitian yang dilakukan. Teori tersebut antara

lain OCR, algoritma K-NN, reduksi K-SVNN, bahasa pemrograman MATLAB 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 ini berisi tentang analisis dan perancangan aplikasi yang dibangun yang meliputi analisis sistem, analisis masalah, analisis metode, analisis kebutuhan non fungsional, pemodelan sistem, serta perancangan dan usulan perbaikan antar muka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi tentang implementasi dan pengujian pada sistem. Tahap implementasi merupakan tahap pembangunan sistem yang sudah dianalisis. Kemudian dilakukan pengimplementasian sistem untuk menguji sistem yang telah dibangun. Metode yang digunakan dalam pengujian sistem adalah pengujian *Black Box* serta pengujian akurasi untuk mengetahui hasil dari klasifikasi menggunakan metode K-NN dan reduksi K-SVNN.

BAB 5 KESIMPULAN DAN SARAN

Bab ini berisi tentang hasil keseluruhan penelitian yang sudah dilakukan. Bagian kesimpulan merupakan hasil dari penelitian yang telah dilakukan dan bagian saran merupakan masukan untuk penelitian dikemudian hari.

