

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Ekstraksi informasi merupakan suatu bidang ilmu pengolahan bahasa alami yang bertujuan untuk mengubah teks tidak terstruktur menjadi informasi dalam bentuk terstruktur[1]. Ekstraksi informasi bisa diterapkan pada berbagai jenis dokumen, dengan cara mendeteksi komponen yang ada pada dokumen tersebut. Dalam penelitian ini ekstraksi informasi akan diterapkan pada sebuah dokumen daftar riwayat hidup. Daftar riwayat hidup merupakan dokumen yang memuat keterangan tentang seseorang[2]. Daftar riwayat hidup atau bisa disebut dengan *Curriculum Vitae* (CV) memiliki jenis-jenis tertentu. Dalam penelitian ini akan digunakan CV lamaran kerja. Untuk melakukan penyaringan informasi dari CV yang ada maka ada suatu divisi *Human Resources Development* atau disingkat *HRD* yang akan melakukan proses penyaringan karyawan dari CV yang ada. Tetapi *HRD* kesulitan melakukan penilaian yang objektif dengan waktu yang singkat dan tidak adanya laporan otomatis seleksi calon karyawan[3]. Dengan demikian, daftar riwayat hidup akan di lakukan proses ekstraksi informasi sehingga menjadi sebuah informasi yang terstruktur. Namun, banyaknya format yang berbeda dari setiap daftar riwayat hidup, membutuhkan suatu metode yang dapat menangani kondisi tersebut. sehingga aturan ekstraksi informasi yang sesuai dengan format dokumen daftar riwayat hidup.

Pada beberapa penelitian mengenai ekstraksi informasi telah dilakukan sebelumnya oleh Agny Ismaya[4]. Pada penelitian tersebut menggunakan sistem berbasis aturan untuk melakukan ekstraksi informasi pada dokumen Laporan Hasil Pemeriksaan (LHP) atas Laporan Keuangan Pemerintah Daerah (LKPD). Permasalahan yang menjadi dasar penurunan akurasi pada penelitian tersebut adalah penentuan pola dokumen. Dari hasil penelitian diketahui bahwa hasil klasifikasi pada data uji mengalami penurunan. Permasalahan yang menjadi dasar penurunan akurasi pada penelitian tersebut adalah penentuan pola dokumen dengan

metode *wordmatch* (kecocokan kata), sehingga apabila terjadi sedikit perbedaan pada dokumen akan terjadi kesalahan dalam klasifikasi. Oleh karena itu, permasalahan tersebut dapat diatasi jika menggunakan *machine learning*.

Pada penelitian sebelumnya, penggunaan *machine learning* di bidang ekstraksi informasi telah dilakukan pada makalah ilmiah[5]. Begitupun pada bidang lainnya, seperti klasifikasi teks. Klasifikasi Berita *Online* Menggunakan Metode *Support Vector Machine* dan *K- Nearest Neighbor*[6] menyatakan bahwa metode yang lebih unggul adalah metode *Support Vector Machine* dengan perolehan akurasi sebesar 93,2%. Sedangkan pada penelitian yang telah dilakukan oleh Mercury dan Adhitya[7] pada Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan metode *Naive Bayes* dan *Support Vector Machine* (SVM), metode *Support Vector Machine* menghasilkan akurasi tertinggi sebesar 99,999%. Sama halnya seperti penelitian pada Analisis Perbandingan Akurasi dalam Identifikasi *Autism* dengan SVM dan *Naive Bayes*[8] metode *Support Vector Machine* (SVM) menghasilkan rata-rata klasifikasi sebesar 93,12% dan lebih unggul dari *Naive Bayes*. Oleh karena itu, algoritma *Support Vector Machine* (SVM) memiliki kemungkinan untuk memperoleh akurasi yang lebih baik apabila dilakukan pada bidang ekstraksi informasi.

Berdasarkan paparan sebelumnya, maka dalam penelitian ini digunakan metode *Support Vector Machine* (SVM) sebagai klasifikasi dalam ekstraksi informasi. Hal ini dimaksudkan untuk mengolah dokumen daftar riwayat hidup menjadi informasi-informasi yang terstruktur dan mengukur tingkat keakuratan algoritma *Support Vector Machine* (SVM) dalam ekstraksi informasi pada daftar riwayat hidup.

1.2 Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini adalah sebagai berikut.

1. Ekstraksi informasi menggunakan *Rule Based* tidak cocok untuk dokumen yang memiliki pola tidak sama sehingga diperlukan *machine learning*.
2. Belum terukurnya kinerja algoritma *Support Vector Machine* (SVM) untuk klasifikasi dalam ekstraksi informasi daftar riwayat hidup.

1.3 Maksud dan Tujuan

Adapun maksud dari penelitian ini adalah mengimplementasikan algoritma *Support Vector Machine* (SVM) kedalam ekstraksi informasi daftar riwayat hidup.

Sedangkan tujuan dari penelitian ini adalah melakukan ekstraksi informasi menggunakan *machine learning* dan mengukur kinerja algoritma *Support Vector Machine* (SVM) dalam ekstraksi informasi daftar riwayat hidup.

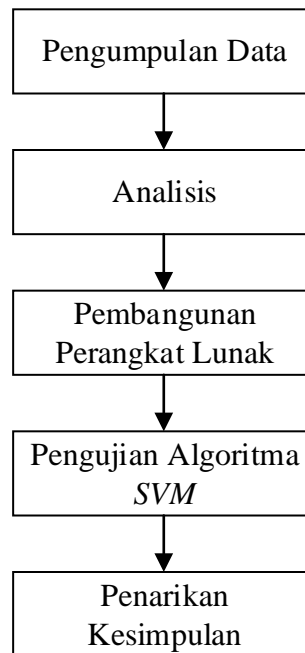
1.4 Batasan Masalah

Adapun batasan masalah dari penelitian ini sebagai berikut.

1. Data *training* dan data *testing* daftar riwayat hidup yang diperoleh dari umum dan internet.
2. Daftar riwayat hidup yang digunakan berupa daftar riwayat hidup untuk lamaran kerja dan berbahasa indonesia.
3. Sumber data yang diambil berupa format *file text* dengan ekstensi pdf (*.pdf).
4. Ekstraksi fitur menggunakan fitur lokal yang dikembangkan menjadi 23 fitur.
5. *Output* yang dihasilkan berupa sub judul, data diri, pendidikan formal, pendidikan non formal, pengalaman kerja, kemampuan dan lain-lain.
6. Pengujian akurasi dengan metode *Confusion Matrix*.

1.5 Metode Penelitian

Dalam penelitian ini, metode penelitian yang digunakan adalah metode deskriptif. metode ini mempelajari masalah-masalah yang ada serta tata cara kerja yang berlaku[9]. Tujuan dari penelitian deskriptif ini adalah untuk membuat deskripsi, gambaran atau lukisan secara sistematis, faktual dan akurat mengenai fakta-fakta, sifat-sifat serta hubungan antar fenomena yang diselidiki[10]. Adapun tahapan dari penelitian ini adalah sebagai berikut.



Gambar 1.1 Alur Tahapan Penelitian

1.5.1 Pengumpulan Data

Pada penelitian ini tahap awal yang dilakukan adalah mengumpulkan data. Adapun metode yang digunakan dalam pengumpulan data adalah sebagai berikut.

1. Studi literatur

Pada penelitian ini dilakukan studi literatur untuk mempelajari teori-teori yang berkaitan dengan metode yang digunakan serta mencari referensi-referensi yang berkaitan dengan penelitian.

2. Pengumpulan data daftar riwayat hidup

Dokumen daftar riwayat hidup yang telah dikumpulkan selanjutnya dijadikan data dalam penelitian ini. dokumen daftar riwayat hidup juga diperoleh dari umum dan internet.

3. Observasi

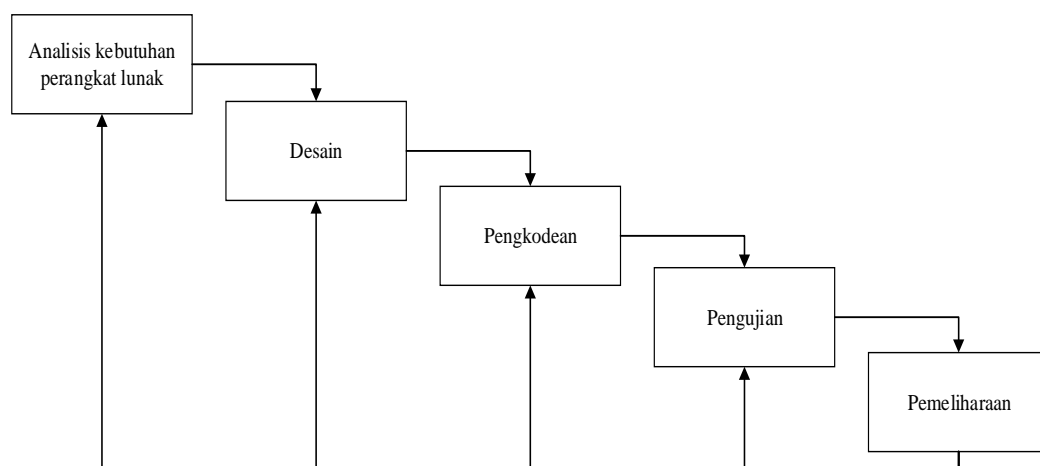
Metode pengumpulan data yang ketiga dalam penelitian ini adalah metode observasi dari pengumpulan data daftar riwayat hidup, yaitu dengan mengamati langsung jenis-jenis daftar riwayat hidup.

1.5.2 Analisis

Tahap selanjutnya yang dilakukan setelah tahap pengumpulan data adalah tahap analisis. tahap analisis dalam penelitian ini meliputi tahap analisis masalah, analisis data masukan, analisis sistem, analisis kebutuhan fungsional serta kebutuhan non-fungsional. analisis sistem pada penelitian ini terbagi menjadi beberapa tahap, yaitu analisis *preprocessing*, pembobotan dan klasifikasi menggunakan *SVM*. Kemudian akan dilanjutkan pada analisis tahapan ekstraksi informasi serta analisis untuk mengukur tingkat akurasi.

1.5.3 Metode Pembangunan Perangkat Lunak

Setelah tahap analisis, tahap selanjutnya yaitu tahap untuk membangun perangkat lunak. Adapun metode pembangunan perangkat lunak yang digunakan dalam penelitian ini adalah metode *waterfall* menurut Sommerville[11]. Metode *waterfall* ini digunakan karena model ini melakukan pendekatan secara sistematis dan berurutan dalam pembangunan perangkat lunak meliputi proses sebagai berikut[11].



Gambar 1.2 Tahap Model *Waterfall*[11]

Adapun penjelasan dari setiap tahapannya adalah sebagai berikut.

1. Analisis kebutuhan perangkat lunak

Langkah ini merupakan analisis terhadap kebutuhan perangkat lunak, dan tahap dalam pengumpulan kebutuhan penelitian secara lengkap tentang ekstraksi informasi pada daftar riwayat hidup. Data yang diambil adalah

data daftar riwayat hidup dalam format .pdf, dan mengumpulkan data-data tambahan.

2. Desain

Setelah kebutuhan data dan pemroses telah dikumpulkan, maka tahap selanjutnya adalah perancangan arsitektur sistem secara keseluruhan berupa struktur data, arsitektur perangkat lunak, representasi antarmuka, dan prosedur pengodean. Tahap ini mentraslasikan kebutuhan perangkat lunak dari tahap analisis kebutuhan ke representasi desain agar dapat diimplementasikan menjadi program pada tahap selanjutnya.

3. Pengkodean

Setelah perancangan sistem secara keseluruhan, maka tahap selanjutnya adalah mengimplementasikan kedalam bahasa pemrograman. Tahap dimana aturan yang telah dibuat akan digunakan dalam menyelesaikan permasalahan dalam ekstraksi informasi pada daftar riwayat hidup.

4. Pengujian

Tahapan ini digunakan untuk mengintegrasikan program dan pengujian tujuan, tahap pengujian terhadap fitur-fitur pada sistem yang telah dibangun dengan menggunakan metode *Black Box*. apabila program yang dibuat berjalan sesuai dengan harapan, dan sesuai dengan aturan yang telah dibuat maka ekstraksi informasi daftar riwayat hidup siap digunakan oleh pengguna sistem.

5. Pemeliharaan

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

1.5.4 Pengujian Algoritma SVM

Tahap selanjutnya adalah tahap melakukan pengujian untuk menghitung keakuratan dari metode SVM yang diimplementasikan untuk ekstraksi informasi daftar riwayat hidup dengan metode *Confusion Matrix*.

1.5.5 Penarikan Kesimpulan

Setelah semua tahap dilakukan dan diselesaikan, tahap terakhir pada penelitian ini adalah melakukan penarikan kesimpulan dari hasil pengujian metode *SVM*.

1.6 Sistematika Penulisan

Adapun sistematika penulisan dalam pembuatan laporan penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab ini menjelaskan mengenai latar belakang dari penelitian yang dilakukan. Selain itu juga menjelaskan identifikasi masalah sesuai dengan latar belakang serta maksud dan tujuannya. Pada bab ini juga menjelaskan metode penelitian yang digunakan serta tahapan dari penelitian.

BAB 2 LANDASAN TEORI

Pada bab ini menjelaskan materi-materi pendukung mengenai daftar riwayat hidup, ekstraksi informasi, metode *SVM* serta hal-hal yang terkait lainnya.

BAB 3 ANALISIS DAN PERANCANGAN

Bab 3 ini mendeskripsikan mengenai analisis masalah, analisis data masukan, analisis sistem, analisis kebutuhan fungsional serta non-fungsional dan analisis perancangan sistem.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini menjelaskan implementasi dari apa yang telah dirancang sebelumnya dan membahas tentang pengujian akurasi algoritma *SVM* pada ekstraksi informasi daftar riwayat hidup dengan menggunakan metode *confusion matrix*.

BAB 5 KESIMPULAN DAN SARAN

Bab terakhir pada penelitian ini menjelaskan mengenai kesimpulan akhir yang didapatkan. Selain itu, pada bab ini juga menjelaskan mengenai saran terhadap sistem yang telah dibangun.

