

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Ekstraksi informasi adalah prosedur untuk mengambil informasi dengan menganalisis hubungan, pola-pola, dan aturan-aturan dari data teks [1]. Dalam penelitian ini ekstraksi informasi diterapkan pada dokumen surat keputusan. Surat keputusan adalah naskah dinas yang memuat kebijakan yang bersifat menetapkan, tidak bersifat mengatur, dan merupakan pelaksanaan kegiatan [2]. Ekstraksi informasi pada surat keputusan bertujuan untuk mengambil informasi yang terdapat pada surat tersebut agar menjadi informasi yang terstruktur. Ekstraksi informasi dapat dilakukan dengan cara klasifikasi. Klasifikasi adalah proses dalam *text mining* yang berfungsi mengelompokkan data ke dalam kelas tertentu yang diberikan berdasarkan sifat dan pola yang terdapat dalam data pembelajaran (*data training*) [3].

Penelitian tentang ekstraksi informasi telah dilakukan oleh Firdamdani [4]. Pada penelitian tersebut ia mengekstraksi informasi pada karya ilmiah menggunakan metode LVQ dengan akurasi 78%. Masalah yang dimiliki pada penelitian tersebut adalah karena kesalahan dalam pembuatan algoritma, salah pengejaan, dan munculnya simbol tidak karuan. Adapun penelitian dari Chandra Ratiwi [5] berhasil mengekstraksi sebuah surat masuk menggunakan metode naive bayes dengan akurasi 96,96%. Walaupun penelitian tersebut memiliki akurasi yang tinggi tetapi penelitian ini tetap memiliki masalah pada bagian konversi di API nya untuk mendapatkan akurasi yang lebih tinggi. Penelitian tentang ekstraksi informasi pun pernah dilakukan oleh Agny Ismaya [6]. Pada penelitian tersebut ia membuat sebuah *tools* dengan algoritma *rule-based classification* untuk melakukan ekstraksi informasi pada dokumen Laporan Hasil Pemeriksaan (LHP) pada Laporan Keuangan Pemerintah Daerah (LKPD). Akan tetapi penelitian tersebut pada hasil klasifikasi data uji mengalami penurunan akurasi. Permasalahan tersebut dikarenakan penelitian tersebut menggunakan metode *wordmatch* (kecocokan kata) pada penentuan pola dokumen, sehingga apabila terjadi sedikit perbedaan pada

dokumen akan terjadi kesalahan dalam klasifikasi tersebut. Permasalahan tersebut kemungkinan dapat diselesaikan bila pada proses klasifikasi dapat menentukan pola dokumen dari segi pembelajaran kemiripan antar dokumennya. Surat keputusan memiliki karakteristik yang berbeda pada setiap instansi dan juga membutuhkan klasifikasi untuk mengenali pola tersebut.

Metode yang akan digunakan untuk proses klasifikasi adalah Support Vector Machine. Pada penelitian "*SVM Based Learning System For Information Extraction*" yang dilakukan Yaoyong LI [7] ia mengekstraksi 3 corpus menggunakan SVM dengan quadratic kernel lain pada penelitian "*Performance Evaluation of SVM-Based Information Extraction using  $\tau$  Margin Values*" yang dilakukan kuspriyanto [8] mengekstraksi informasi *corpus* pasar sayur menggunakan SVM dengan quadratic kernel. Dikatakan pada penelitian tersebut bahwa SVM memiliki performansi yang cukup tinggi dan bila semakin banyak fitur maka akurasi akan semakin tinggi. Namun SVM ini belum diterapkan pada ekstraksi informasi dokumen surat keputusan, sehingga belum diketahui bila diterapkan akan memiliki akurasi yang baik atau tidak.

Berdasarkan paparan sebelumnya, maka dalam penelitian ini digunakan metode SVM sebagai klasifikasi dalam ekstraksi informasi. Hal ini bermaksud untuk mengolah dokumen surat keputusan menjadi informasi yang terstruktur dan mengukur tingkat akurasi metode SVM dalam ekstraksi informasi pada dokumen surat keputusan.

## **1.2 Rumusan 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 SVM dikarenakan ia memiliki hasil yang baik dalam ekstraksi informasi.

### **1.3 Maksud dan Tujuan**

Berdasarkan identifikasi masalah tersebut maka dapat disusun maksud dan tujuan sebagai berikut:

#### **1.3.1 Maksud**

Berdasarkan identifikasi masalah tersebut maka dapat disusun maksud dari penelitian ini adalah:

1. Membuat sistem ekstraksi informasi pada surat.
2. Mengimplementasi SVM dalam ekstraksi informasi pada dokumen surat keputusan.

#### **1.3.2 Tujuan**

Berdasarkan identifikasi masalah tersebut maka dapat disusun tujuan dari penelitian ini adalah:

1. Mengetahui informasi yang terstruktur pada dokumen surat keputusan setelah diklasifikasi.
2. Mengukur performa SVM dalam ekstraksi informasi pada dokumen surat keputusan dengan baik.

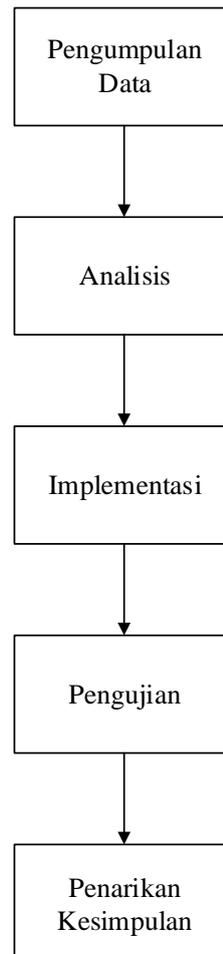
### **1.4 Batasan Masalah**

Agar penelitian dapat dilakukan lebih fokus dalam memandang permasalahannya maka perlu dibatasi variabelnya. Batasan permasalahannya sebagai berikut:

1. Data yang dipakai dalam bentuk teks dengan ekstensi pdf (\*.pdf).
2. Data training dan data testing diperoleh dari internet ataupun umum.
3. Data merupakan surat keputusan yang ditujukan hanya kepada satu orang.
4. Data minimal terdiri dari nomor surat, nama penerima surat keputusan dan waktu ditetapkannya surat keputusan
5. Data harus merupakan gambar dengan kualitas yang baik.

## 1.5 Metode Penelitian

Metode Penelitian yang digunakan adalah metode deskriptif. Metode deskriptif merupakan metode yang menggambarkan fakta-fakta dan informasi dalam situasi atau kejadian sekarang secara sistematis, faktual dan akurat. Adapun gambaran alur dalam penelitian ini dapat dilihat pada Gambar 1.1.



**Gambar 1.1 Alur Penelitian**

### 1.5.1 Tahap Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Studi Literatur

Tahap studi literatur dilakukan dengan cara mempelajari sumber seperti hasil penelitian, buku referensi, dan bacaan-bacaan yang berkaitan dengan

penelitian yang berhubungan dengan administrasi surat, *text mining*, dan support vector machine.

b. Observasi

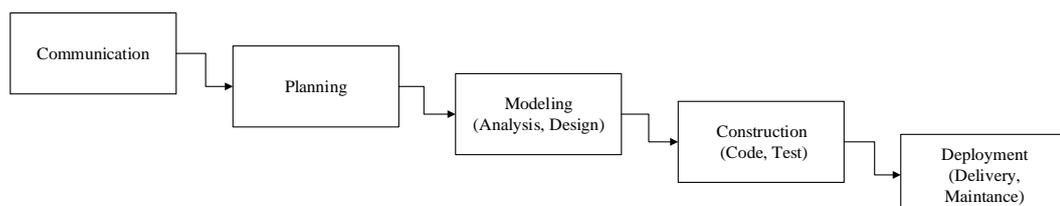
Teknik pengumpulan data observasi dilakukan dengan mengamati langsung surat-surat keputusan yang berlaku baik langsung dari suatu instansi ataupun yang ada pada internet

### 1.5.2 Tahap Analisis

Setelah dilakukan pengumpulan data dan studi literatur, selanjutnya dilakukan analisis dan interpretasi data serta literatur untuk mendapatkan secara rinci mengenai tahapan-tahapan apa saja yang akan dilakukan. Analisis terbagi dalam beberapa tahapan, yaitu *preprocessing* dan klasifikasi. Kemudian akan dilanjutkan ke tahapan ekstraksi informasi serta analisis untuk mengukur tingkat akurasi. Dilakukan juga perhitungan secara manual bila diperlukan untuk tiap tahapan yang bertujuan sebagai simulasi yang nantinya akan diterapkan pada tahap selanjutnya yaitu implementasi.

### 1.5.3 Tahap Implementasi

Metode *waterfall* adalah metode yang melakukan pendekatan secara sistematis dan berurutan. Metode ini disebut dengan *waterfall* karena tahap demi tahap yang dilalui harus menunggu selesainya tahap sebelumnya dan berjalan berurutan [9]. Tahapan dalam metode *waterfall* menurut Pressman dapat dilihat pada gambar 1.2.



**Gambar 1.2 Model Metode Waterfall [9]**

a. *Communication*

Komunikasi dalam penelitian ini adalah mengumpulkan data-data yang diperlukan, serta membantu mendefinisikan fitur dan fungsi sistem ekstraksi informasi yang akan dibangun. Jurnal dan artikel yang berhubungan dengan judul penelitian digunakan pula sebagai sumber pengumpulan data-data.

b. *Planning*

Setelah dilakukan observasi, pengumpulan data dan literatur, berikutnya adalah tahapan perencanaan. Tahapan ini menjelaskan tentang sumber daya yang diperlukan dalam membuat sistem ekstraksi informasi, hasil yang diinginkan, dan jadwal proses pembangunan sistem dalam penelitian.

c. *Analysis dan Design*

Setelah dilakukan perencanaan, data dokumen surat keputusan hasil pengumpulan dan observasi dianalisis sesuai literatur yang telah dipelajari. Hasil analisis dan interpretasi data tersebut kemudian dianalisis kebutuhan fungsional dan non fungsional untuk pembangunan perangkat lunak. Dilakukan pula penyusunan desain awal atau gambaran sistem yang akan dibangun.

d. *Coding dan Pengujian*

Setelah analisis dan desain selesai, kemudian dilakukan implementasi atau penulisan *source code* program berupa web dari hasil gambaran awal sistem. Setelah implementasi akan dilakukan pengujian semua fungsi terhadap hasil *source code* program dengan metode *black box*.

e. *Deployment*

Tahapan ini adalah pemeliharaan software dalam jangka waktu yang berkala. Namun dalam penelitian ini tahapan ini tidak digunakan karena tujuan dari sistem yang dibangun dalam penelitian adalah sebagai alat untuk dilakukannya pengujian teori bukan sebagai perangkat lunak untuk digunakan secara berkala oleh user atau instansi.

#### **1.5.4 Tahap Pengujian**

Fase ini adalah fase untuk melakukan pengujian dan menghitung keakuratan dari metode Support Vector Machine yang diimplementasikan untuk ekstraksi informasi dokumen surat keputusan dengan metode *confusion matrix*.

#### **1.5.5 Tahap Penarikan Kesimpulan**

Fase ini akan menyajikan kesimpulan dari hasil pengujian metode Support Vector Machine sesuai dengan tujuan penelitian.

### **1.6 Sistematika Penulisan**

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

#### **1.6.1 BAB 1 Pendahuluan**

Pada bab ini akan dibahas mengenai latar belakang masalah, rumusan masalah, maksud dan tujuan, batasan masalah, metode penelitian, dan sistematika penulisan laporan.

#### **1.6.2 BAB 2 Landasan Teori**

Pada bab ini dijelaskan tentang teori-teori yang digunakan dalam menganalisis permasalahan yang ada seperti definisi SVM, macam-macam jenis surat, dan pembahasan alat yang dipakai dalam penelitian.

#### **1.6.3 BAB 3 Analisis Dan Perancangan**

Pada bab ini berisi tentang analisis masalah, data masukan, dan perancangan sistem serta batasan dari perancangan program dalam tahapan *preprocessing* dan proses klasifikasi menggunakan SVM.

#### **1.6.4 BAB 4 Implementasi Dan Pengujian**

Pada bab ini berisi tentang pengujian akurasi algoritma Support Vector Machine (SVM) pada ekstraksi informasi dokumen surat keputusan dengan menggunakan metode *confusion matrix*.

#### **1.6.5 BAB 5 Kesimpulan Dan Saran**

Pada bab ini merupakan bab penutup yang mengemukakan kesimpulan dan saran untuk penelitian yang selanjutnya.