

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Ekstraksi informasi adalah cabang ruang lingkup dari *text mining* yang bertujuan untuk mengubah hasil proses *text mining* menjadi akar yang sama dengan dunia data yang terstruktur dalam *data mining*[1]. Dalam penelitian ini ekstraksi informasi diterapkan pada dokumen surat masuk. Surat masuk adalah semua surat dinas yang diterima[2]. Ekstraksi informasi pada surat masuk bertujuan untuk mengambil informasi yang terdapat pada surat tersebut agar menjadi informasi yang terstruktur dengan menggunakan *text mining* dan klasifikasi teks. *Text mining* adalah prosedur untuk mengambil informasi dengan menganalisis hubungan, pola-pola, dan aturan-aturan dari data teks[1], sementara klasifikasi adalah suatu 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 mengenai ekstraksi informasi telah dilakukan sebelumnya oleh Agny Ismaya[4]. Pada penelitian tersebut dibuat sebuah *tools* dengan *rule-based classification* untuk melakukan ekstraksi informasi pada dokumen Laporan Hasil Pemeriksaan (LHP) atas Laporan Keuangan Pemerintah Daerah (LKPD)[4]. 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. Permasalahan ini kemungkinan dapat diselesaikan jika metode klasifikasi dapat menentukan pola dokumen dari segi pembelajaran kemiripan antar dokumennya atau *machine learning*.

Dokumen yang digunakan dalam ekstraksi informasi pada penelitian ini adalah dokumen surat masuk. Hermawan S. melakukan penelitian administrasi di Kel. Jeruk Kab Sragen, dengan pembuatan aplikasi yang bertujuan untuk menyimpan data dan sebagai sarana temu balik apabila berkas surat tersebut

digunakan kembali[5]. Namun, dalam hasil penelitian tersebut penyimpanan data surat masuk yang diantaranya terdiri dari pengirim, nomor surat, tanggal surat, dan perihal masih diinputkan secara manual sehingga prosesnya masih sama dengan pencatatan data surat masuk ke buku agenda atau pengetikan manual ke data arsip komputer. Dengan ekstraksi informasi pada dokumen surat masuk maka pengambilan informasi dapat dilakukan secara otomatis sehingga akan meringkas proses input data secara manual. Namun ide ekstraksi informasi menggunakan *rule-based* kurang mampu menangani karakteristik data yang tidak identik seperti surat masuk, sehingga tetap dibutuhkannya metode lain untuk dapat menangani klasifikasi karakteristik data yang tidak identik seperti dokumen surat masuk.

Metode yang akan digunakan untuk proses klasifikasi adalah *Fuzzy K-Nearest Neighbour*. Pada penelitian “*Web Document Classification Based on Fuzzy k-NN Algorithm*” yang dilakukan oleh Zhang dkk[6] serta pada penelitian “Penerapan *Fuzzy K-NN* Dalam Menentukan Status Gizi Balita” yang dilakukan Satria N. dkk[7], dikatakan bahwa kelebihan dari metode *fuzzy K-NN* terletak pada tingkat akurasi yang dihasilkan lebih baik daripada metode klasifikasi lain seperti *Support Vektor Machine (SVM)*, dan *K-Nearest Neighbour (K-NN)*. Namun *Fuzzy K-NN* belum diterapkan dalam ekstraksi informasi dokumen, sehingga belum diketahui apakah bila diterapkan tingkat akurasi *Fuzzy K-NN* tetap baik atau tidak.

Berdasarkan paparan sebelumnya, maka dalam penelitian ini digunakan metode *Fuzzy K-Nearest Neighbour* sebagai klasifikasi dalam ekstraksi informasi. Hal ini dimaksudkan untuk mengolah dokumen surat masuk menjadi informasi-informasi yang terstruktur dan mengukur tingkat keakuratan algoritma *Fuzzy K-NN* dalam ekstraksi informasi pada dokumen surat masuk.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang tersebut terdapat identifikasi masalah sebagai berikut :

1. Karakteristik data dokumen surat masuk yang tidak identik sehingga membutuhkan metode klasifikasi dalam ekstraksi informasi yang

mampu menangani karakteristik data tersebut.

2. Belum terukurnya tingkat akurasi dari metode klasifikasi *Fuzzy K-nearest Neighbour* (*Fuzzy K-NN*) bila diterapkan dalam ekstraksi informasi.

### 1.3 Maksud dan Tujuan

Maksud dari penelitian ini adalah untuk melakukan ekstraksi informasi pada dokumen surat masuk menggunakan metode *Fuzzy k-Nearest Neighbour*.

Adapun tujuan dari penelitian ini adalah :

1. Menerapkan metode klasifikasi *Fuzzy K-Nearest Neighbour* (*Fuzzy K-NN*) dalam ekstraksi informasi.
2. Mengukur tingkat akurasi dari implementasi metode *Fuzzy K-nearest Neighbour* (*Fuzzy K-NN*) dalam ekstraksi informasi dokumen surat masuk.

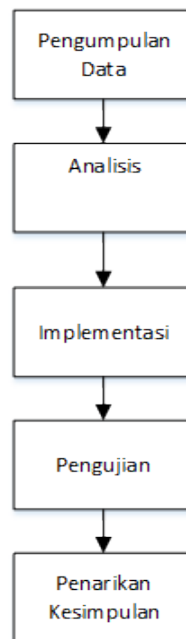
### 1.4 Batasan Masalah

Adapun batasan masalah dari penelitian tersebut sebagai berikut :

1. Data *training* dan data *testing* diperoleh dari surat dinas resmi yang didapat secara umum.
2. Data yang diambil sesuai atau menyerupai dengan aturan penulisan surat dinas resmi Indonesia (Instansi Pemerintah/Non-Instansi Pemerintah) menurut Pedoman Tata Naskah Dinas Instansi Pemerintah.
3. Data yang diambil sedikitnya memiliki kop surat, tanggal surat, nomor surat, dan perihal surat.
4. Sumber data diambil disimpan dalam format *file text* dengan ekstensi pdf (\*.pdf).
5. Ekstraksi fitur menggunakan fitur identitas dan fitur lokal yang dikembangkan menjadi 11 parameter pembobotan.[8][9]
6. *Euclidean distance* digunakan sebagai metode penghitungan jarak antara data *testing* dan data *training*.[10]

## 1.5 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini yaitu metode penelitian deskriptif. Tujuan dari penelitian deskriptif ini adalah untuk membuat deskripsi sistematis, dan penjelasan mengenai hal yang akan diteliti serta menghubungkan dengan teori-teori atau hipotesa yang sudah ada. Dengan metode penelitian deskriptif yang pertama dilakukan adalah pengumpulan data dan referensi serta observasi pada surat dinas yang berlaku. Kemudian dilakukan analisis yang menghasilkan desain sistem. Setelah itu dilakukan implementasi dan pengujian sistem, dan pada akhirnya akan dihitung tingkat akurasi juga pengambilan kesimpulan dari penelitian. Adapun gambaran alur dalam penelitian ini dapat dilihat pada Gambar 1.1.



**Gambar 1.1** Gambaran Alur Tahapan 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 kepustakaan diantaranya hasil penelitian, buku referensi, dan bacaan-bacaan yang berkaitan dengan penelitian terutama penelitian yang

berhubungan dengan administrasi surat dan *text mining*.

b. Observasi

Teknik pengumpulan data observasi dilakukan dengan mengamati langsung surat-surat resmi yang berlaku baik langsung dari suatu instansi maupun surat edaran atau undangan resmi yang diunggah ke internet.

c. Pengumpulan Data Dokumen Surat Masuk

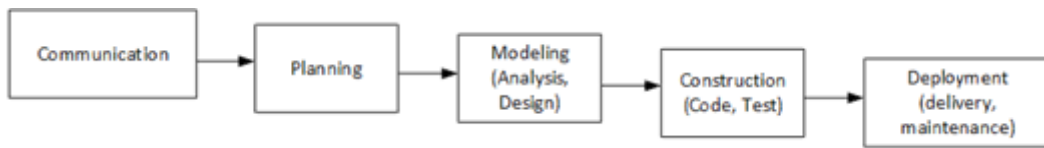
Dokumen surat masuk yang telah di observasi dan sesuai dengan literatur kemudian dikumpulkan untuk dijadikan data dalam penelitian ini. Surat diperoleh dari instansi langsung seperti UNPAD dan Pemerintah Kab. Pangandaran, serta surat yang dinas resmi yang diunggah ke internet seperti dari Pengadilan Tinggi dan Kemenristekdikti.

### 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 menjadi dalam beberapa tahapan, yaitu tahapan *preprocessing* dan tahapan klasifikasi. Kemudian akan dilanjutkan pada analisis tahapan ekstraksi informasi serta analisis untuk mengukur tingkat akurasi. Dilakukan pula perhitungan manual bila diperlukan untuk masing-masing tahapan yang bertujuan sebagai simulasi yang nantinya akan diterapkan pada tahap selanjutnya yakni implementasi.

### 1.5.3 Metode Pembangunan Perangkat Lunak (Implementasi)

Metode *waterfall* adalah metode yang melakukan pendekatan secara sistematis dan berurutan. Metode ini disebut dengan *waterfall* karena tahapan demi tahapan yang dilalui harus menunggu selesainya tahapan sebelumnya dan berjalan berurutan[11]. Tahapan dalam metode *waterfall* menurut Pressman dapat dilihat pada Gambar 1.2.



**Gambar 1.2 Gambar Tahap Model *Waterfall*[11]**

*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, artikel, dan internet 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, hasil yang diinginkan, dan jadwal proses pembangunan sistem dalam penelitian.

*c. Analysis dan Design*

Setelah dilakukan perencanaan, data dokumen surat hasil pengumpulan dan observasi dianalisis sesuai literatur yang telah dipelajari. Hasil analisis dan interpretasi data tersebut kemudian disusun analisis 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 dari hasil gambaran awal sistem. Setelah implementasi akan dilakukan pengujian semua fungsi terhadap hasil *source code* program.

*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 dengan metode *black box* dan menghitung keakuratan dari metode *Fuzzy K-nearest Neighbour* yang diimplementasikan untuk ekstraksi informasi dokumen surat masuk dengan metode *Confusion Matrix*, yang memiliki 4 keluaran, yaitu: *recall*, *precision*, *f1-score* dan *acuraccy*[3].

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

Fase ini akan menyajikan kesimpulan dari hasil pengujian metode *Fuzzy K-Nearest Neighbour* sesuai dengan tujuan penelitian.

### **1.6 Sistematika Penulisan**

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

#### **BAB 1 PENDAHULUAN**

Bab ini berisi latar belakang, identifikasi masalah, maksud dan tujuan, batasan masalah, metode penelitian dan sistematika penulisan.

#### **BAB 2 LANDASAN TEORI**

Bab ini berisi uraian yang berkaitan tentang surat, format surat, *text preprocessing*, serta teori yang berkaitan mengenai Logika *Fuzzy*, Algoritma *K-Nearest Neighbour*, dan *Fuzzy K-Nearest Neighbour*.

#### **BAB 3 ANALISIS DAN PERANCANGAN**

Bab ini membahas mengenai analisis masalah, analisis metode yang digunakan, analisis penyelesaian masalah, analisis simulasi (perhitungan metode *Fuzzy K-Nearest Neighbour*) klasifikasi dan analisis perancangan sistem.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini membahas tentang pengujian akurasi algoritma *Fuzzy K-nearest Neighbour* (*Fuzzy K-NN*) pada ekstraksi informasi dokumen surat masuk dengan menggunakan sistem yang telah dibangun dengan menggunakan metode *confusion matrix*.

## **BAB 5 KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dan saran dari penelitian yang dilakukan.