# BAB 1
# PENDAHULUAN

## 1.1 Latar Belakang Masalah

Surat merupakan salah satu alat yang digunakan untuk berkomunikasi antara satu pihak dan pihak lainnya untuk menyampaikan suatu informasi yang tertulis. Setiap surat memiliki nilai informasi yang sudah pasti seperti nama instansi pengirim, alamat, tanggal, perihal dan sebagainya. Untuk mendapatkan informasi tersebut maka timbul kebutuhan untuk mengekstraksi informasi secara otomatis dari data yang berupa teks pada surat tersebut agar dapat disimpan dalam bentuk data [1].

Ekstraksi informasi adalah pencarian otomatis pada informasi yang terstruktur seperti entitas, hubungan antar entitas, dan atribut yang menggambarkan entitas dari sumber yang tidak terstruktur[2]. Pencarian dari suatu ekstraksi informasi merupakan sebuah sistem yang berfungsi untuk mencari data spesifik dalam *Natural Language Text.* Dalam sebuah *template* formulir atau tabel serta komponen – komponen teks yang ada diambil datanya untuk diekstraksi menjadi data yang memiliki kegunaan [3]. Ekstraksi informasi pada teks berpola sebelumnya pernah diteliti oleh Dimas Mustaqwa dengan judul "Implementasi Ekstrasi Informasi Pada Dokumen Skripsi Menggunakan Metode *Ruled Based*" akan tetapi walaupun tingkat akurasi yang dihasilkan oleh metode tersebut sudah 100% namun metode tersebut harus memiliki pola yang sama persis dengan dataset sehingga tidak cocok dengan teks yang tidak selalu memiliki pola yang sama [4]. Berdasarkan masalah pada *rule based* tersebut maka menurut Marina Dorash bahwa solusi yang dapat dilakukan adalah dengan menggunakan *machine* learning [5]. Salah satu metode dari *machine learning* adalah *Naïve Bayes. Naïve Bayes* memiliki beberapa keunggulan dan kelebihan dari algoritma lainnya yaitu sederhana, cepat dan memiliki tingkat akurasi yang tinggi [6].

Ekstraksi informasi pernah dilakukan oleh Oman Sumantri dalam penelitiannya bahwa Naïve Bayes memiliki tingkat akurasi 14% lebih baik dibandingkan dengan SVM [7]. Hal tersebut diperkuat bahwa berdasarkan

penelitian yang dibuktikan oleh Xhemali, Hinde dan Stone dalam penelitiannya mengatakan bahwa Naïve Bayes Classifier memiliki tingkat akurasi yang lebih baik dibandingkan Decision Tree dan Neural Network[8].

Berdasarkan hal diatas maka pada penelitian ini *Naïve Bayes* digunakan sebagai classifier untuk mengklasifikasikan teks yang terkandung dalam surat dengan mempelajari dari informasi surat, kemudian sistem mengekstraksi informasi dari hasil klasifikasi tersebut menjadi informasi yang dapat diolah kembali ataupun digunakan dengan kepentingan lainnya.

## 1.2 Identifikasi Masalah

Berdasarkan masalah yang sudah dijelaskan pada latar belakang maka identifikasi masalah meliputi hal berikut yaitu :

1) Ekstraksi informasi menggunakan *Rule Based* tidak cocok untuk surat masuk yang memiliki pola tidak selalu sama sehingga diperlukan *machine learning.*
2) Seberapa baik kinerja *Naïve Bayes* untuk klasifikasi text mining dalam ekstraksi informasi.

## 1.3 Maksud dan Tujuan

Berdasarkan permasalahan yang diteliti, maka maksud dari penulisan tugas mengimplementasikan algoritma *Naïve Bayes* kedalam ekstraksi informasi. Adapun tujuan dari tugas akhir ini adalah dapat mengukur kinerja *Naïve Bayes* dalam ekstraksi informasi surat masuk.

## 1.4 Batasan Masalah

Adapun batasan masalah dari penelitian tersebut sebagai berikut :

1. Data *Training* dan *Testing* diperoleh dari surat dinas resmi yang didapat dari umum dan internet.
2. Data yang diambil sesuai dengan aturan – aturan penulisan surat dinas resmi Indonesia (Instansi Pemerintah/ Non – Instansi Pemerintah).

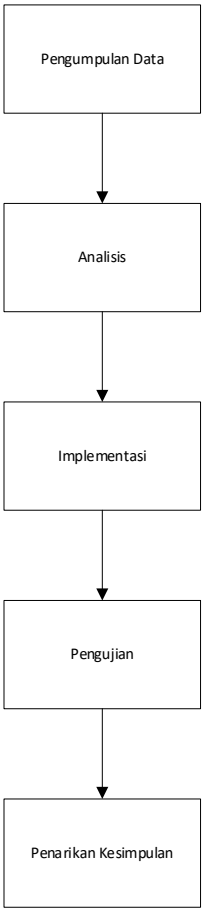
3. Data yang diambil sedikitnya memiliki nama instansi, tanggal surat, nomor surat dan perihal surat.
4. Sumber data diambil dan disimpan dalan bentuk *file text* dengan ekstensi pdf (*.pdf).
5. Data surat harus dalam kualitas gambar yang baik agar dapat terbaca saat konversi.

## 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.

### 1.5.1 Alur Penelitian

Alur penelitian dapat dilihat pada gambar 1.1 :




**Gambar 1.1 Alur Penelitian**

### 1.5.2 Metode Pengumpulan Data

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

a. Studi Literatur

   Pengumpulan data dengan cara mengumpulkan literatur, jurnal, paper dan serta bacaan yang berkaitan dengan judul penelitian.

b. Obeservasi.

   Metode pengumpulan data dengan cara mengadakan penelitian dan peninjauan langsung terhadap masalah yang diambil.

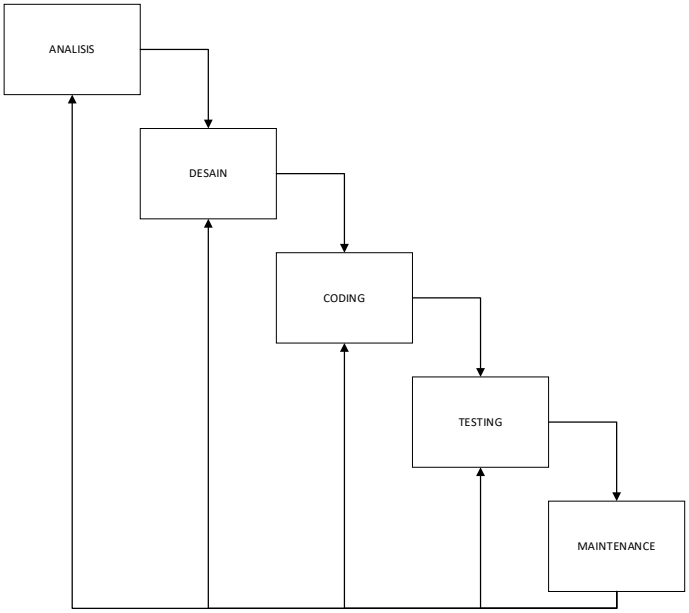
### 1.5.3 Metode Tahapan Analisis

Berdasarkan pengkajian dan evaluasi setelah membaca beberapa sumber seperti jurnal, dan buku – buku, Tahapan analisis pada penelitian ini adalah :

1. Menganalisis Surat, yaitu menganalisis pola surat dan bentuk surat, surat yang dianalisis adalah surat masuk dengan format resmi dengan tahapan *preprocessing* dengan ekstraksi fitur.
2. Menganalisis tahapan klasifikasi, yaitu melakukan perhitungan menghitung data latih sehingga menghasilkan nilai yang akan digunakaan pada saat tahap pengujian, tahapan yang digunakan dalam tahapan pelatihan ini adalah dengan menggunakan metode *naïve bayes,* hasil dari tahapan analisis pelatihan ini adalah nilai parameter *naïve bayes* terhadap data baru yang akan diuji.
3. Menganalisis tahapan pengujian, yaitu bagaimana aplikasi setelah diterapkannya dari penganalisisan dari metode *naïve bayes* untuk mengekstraksi informasi dari inputan baru. Hasil akhir dari tahapan ini adalah aplikasi yang dapat mengekstraksi informasi.
4. Menganalisis data keluaran atau tabel ekstraksi, yaitu melakukan analisis terhadap nilai akhir yang didapat dari tahap pengujian, apakah memiliki akurasi yang baik atau tidak.

### 1.5.4 Implementasi

Implementasi menggunakan model *Waterfall*. Model *Waterfall* adalah model pembangunan yang bertahap secara sequensial mulai dari tahapan awal sampai akhir yang saling berketerkaitan. Alur dari metode *Waterfall* dapat dilihat pada gambar 1.2 :




**Gambar 1.2 Metode Waterfall** [9]

1. *Analisis*

   Analisa yang dilakukan pada tahap ini adalah menganalisis segala hal yang diperlukan dalam pelaksanaan proyek pembuatan perangkat lunak .

2. *Design*

   Tahap dimana dilakukan penuangan pikiran dan perancangan sistem yang sesuai dengan solusi kedalam bentuk yang lebih mudah dimengerti oleh user.

3. *Coding*

   Tahap penulisan kode program atau koding merupakan penerjemahan *design* dalam bahasa yang dikenali oleh komputer yang berisikan pemecahan masalah.

4. *Testing*

   Tahap ini dilakukan apakah perangkat lunak sudah dapat memecahkan masalah sesuai dengan apa yang telah dirancang.

5. *Maintenance*

   Merupakan tahap akhir dari perancangan perangkat lunak. Tahap ini dapat mengalami perubahan – perubahan d

   an penambahan sesuai dari permintaan user atau kebutuhan sistem.

### 1.5.4 Metode Pengujian

Dalam tahapan ini akan dilakukan pengujian akurasi dari metode algoritma *Naïve Bayes* yang sudah diimplementasikan ke sistem ekstraksi informasi surat masuk dengan metode Confusion Matrix, dengan 4 keluaran, yaitu *recall, precision, accuracy* dan *error code* [10].

## 1.6 Penarikan Kesimpulan

Merupakan tahapan yang akan menyajikan hasi keseluruhan dari hasil penelitian dan nilai akurasi yang dihasilkan dari metode Algoritma *Naïve Bayes* serta kesimpulan terkait pengujian yang telah dilakukan

## 1.7 Sistematika Penulisan

Sistematika penulisan skripsi yang akan dibuat meliputi beberapa bagian diantaranya sebagai berikut:

**BAB 1 Pendahuluan**

Bab ini berisikan tentang latar belakang, batasan masalah, maksud dan tujuan penelitian, metodologi penelitian dan sistematika dari skripsi yang di buat.

**BAB 2 Landasan Teori**

Bab ini berisikan tentang teori – teori pendukung yang berkaitan tentang judul penelitian meliputi *Artificial Intelligence*, *Naïve Bayes*, *Data Mining* dan *Information Extraction, PHP*

.**BAB 3 Analisis dan Perancangan Sistem**

Bab ini berisi tentang analisis dan perancangan aplikasi yang dibangun yang meliputi analisis masalah, analisis data masukan, analisis sistem, analisis *preprocessing*, analisis algoritma *Naïve Bayes*, analisis non fungsional dan fungsional.

**BAB 4 Implementasi dan Pengujian**

Bab ini berisi tentang gambar dari hasil implemetasi dan pengujian pada sistem yang kemudian di uji mulai scanning untuk mendapatkan data training, serta data latih, proses pengklasifikasian, dan pengumpulan data ekstraksi yang sudah diklasifikasi oleh *Naïve Bayes* dengan metode pengujian *confusion matrix* pada sistem aplikasi yang telah diprogram untuk menguji akurasi dari pengektraksian informasi dari surat masuk.

**BAB 5 Kesimpulan dan Saran**

Bab ini berisi tentang hasil keseluruhan penelitian yang sudah dilakukan apakah sudah tercapai atau belum dan di tunjukan untuk pengembangan di kemudian hari.