

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dengan semakin banyak orang yang terhubung ke Internet, *electronic mail* (*e-mail*) menjadi salah satu bentuk komunikasi yang paling cepat dan ekonomis. *e-mail* yang sangat mudah dan cepat dalam penggunaannya, *e-mail* selain dimanfaatkan untuk berhubungan antara teman biasa maupun rekan bisnis tetapi *e-mail* juga digunakan sebagai salah satu media penyampaian berita dalam bidang *electronic commerce*. Sayangnya dari keuntungan yang membuat *e-mail* sangat populer di masyarakat, pebisnis juga turut memanfaatkan keuntungan yang dimiliki oleh *e-mail*, dengan cara melakukan pengiriman banyak *e-mail* ke *mail box* secara acak, yang berisi tentang barang yang akan dijual, cara menjadi kaya raya dengan cepat atau pun bagaimana mengakses situs - situs dewasa.

Dengan semakin banyaknya orang yang mengirim iklannya lewat Internet dan semakin banyaknya penyedia layanan *e-mail* di Internet, jumlah dari *e-mail spam* semakin bertambah banyak. Sebagai hasilnya banyak pembaca *e-mail* yang harus menghabiskan banyak waktunya untuk melakukan pemilihan secara manual terhadap *e-mail* tersebut. Tetapi *e-mail spam* memiliki kerugian yang lain seperti bahwa *e-mail spam* yang bisanya dikirim secara masal akan memenuhi media penyimpanan pada *e-mail server*, khususnya pada server dengan account yang berjumlah banyak.

Untuk menanggulangi masalah tersebut diatas diperlukan sebuah metode yang secara otomatis dapat melakukan filter atau seleksi *e-mail* yang dapat memisahkan antara *spam e-mail* dan *legitimate e-mail* (*e-mail* selain *spam e-mail*). Sistem Spam Filter harus secara otomatis dapat mengenali perubahan yang tentu saja akan selalu dilakukan oleh para pembuat *spam e-mail*, agar *e-mail* yang dibuat selalu dapat melewati sistem sejenis.

*Naive Bayes* merupakan salah satu *Probabilistic Classifier* yang dapat digunakan untuk dapat mendeteksi *spam e-mail*. *Naive bayes* yang berdasarkan

pada teorema bayes, menggunakan asumsi bahwa masing – masing variabel yang ada didalamnya bersifat tidak saling tergantung. Metode ini didasarkan pada berapa banyak *e-mail* yang mengandung sebuah kata atau gabungan kata ataupun gabungan karakter tertentu, yang digolongkan sebagai *spam e-mail* atau *Legitimate e-mail*. Dengan data latih yang tepat metode ini dapat dengan akurat melakukan filterisasi *e-mail*.

Dengan menggunakan filter yang otomatis ini diharapkan *spam e-mail* dapat dipisahkan dengan *legitimate e-mail*. Yang pada akhirnya akan meringankan beban *e-mail* server sendiri ataupun pengguna *e-mail* server tersebut.

## 1.2 Rumusan Masalah

Permasalahan dalam Spam Filter adalah

- Bagaimana menerapkan dan menggabungkan *Naive Bayesian Bernoulli* dan *Domain Specific Feature* yang digunakan sebagai Spam Filter.
- Bagaimana menentukan daftar *Domain Specific Feature*.
- Bagaimana menentukan daftar kata bentukan.
- Bagaimana menentukan *treshold* yang digunakan batas penentu sebuah *e-mail* dikatakan sebagai *e-mail spam*.

## 1.3 Batasan Masalah

- Aplikasi ini menggunakan inputan *e-mail* yang berupa teks bukan gambar.
- Menggunakan *Naive Bayesian Classifier Bernoulli* sebagai metode klasifikasi.
- Aplikasi ini berdiri sendiri dan tidak menempel pada *e-mail* server.
- *E-mail* yang digunakan sudah dalam bentuk file yang disimpan dalam folder tersendiri.
- Menggunakan inputan *e-mail* dalam bahasa Inggris.

## 1.4 Tujuan Tugas Akhir

- Membuat Aplikasi *spam* Email Filtering.

- Menganalisis keakuratan hasil klasifikasi yang dihasilkan oleh Aplikasi Spam Filter menggunakan *Naive Bayes*.
- Menganalisis pengaruh dari penambahan inputan berdasarkan kata bentukan dan *Domain Specific Feature* pada *Naive Bayes*.
- Menganalisis pengaruh dari perubahan nilai *threshol*d pada aplikasi.

### 1.5 Metodologi Penelitian

- Studi literatur.  
Mengumpulkan bahan-bahan referensi Tugas Akhir yang meliputi topik-topik *Naive bayesian classifier*, *Feature Selection*, rekayasa perangkat lunak, dan topik lainnya yang mendukung penyusunan Tugas Akhir ini. Sumber referensi berupa buku, Tugas Akhir, jurnal, slide presentasi dan *e-book*. Referensi ini akan digunakan sebagai dasar teori penyusunan Tugas Akhir.
- Pengumpulan Data  
Data yang dibutuhkan dalam penyusunan Tugas Akhir ini adalah *e-mail* yang akan digunakan dalam proses pelatihan untuk aplikasi. Data tersebut diperoleh dari *e-mail* yang terdapat pada *inbox outlook express* user tertentu. Dari data yang diperoleh maka akan didesain Perangkat Lunak yang nantinya akan digunakan untuk melakukan filter terhadap *spam e-mail*.
- Pengembangan Perangkat Lunak yang meliputi kegiatan:
  1. Analisa Kebutuhan Sistem  
Pembuatan perancangan perangkat lunak
  2. Pengkodean  
Implementasi perancangan menjadi perangkat lunak
  3. Pengujian Perangkat Lunak  
Perangkat lunak diuji dengan menggunakan data dengan beberapa kasus uji.
  4. Analisa hasil pengujian  
Hasil pengujian dianalisis untuk proses *debugging*.

5. Evaluasi Perangkat Lunak

Perangkat lunak yang dihasilkan dievaluasi berdasarkan data yang diperoleh dari hasil pengujian.

6. Pembuatan Dokumentasi Perangkat Lunak.

## 1.6 Sistematika Penulisan

### **BAB I       PENDAHULUAN**

Berisi tentang latar belakang, perumusan masalah, batasan masalah, metode penyelesaian masalah dan sistematika penulisan.

### **BAB II       LANDASAN TEORI**

Berisi uraian beberapa landasan teori singkat mengenai *Naïve Bayes*, *Email Filtering*, *spam*, *Domain Specific Feature*, *Feature Selection* dan *Zip’f Law*.

### **BAB III      ANALISIS DAN DESAIN SISTEM**

Berisi mengenai pengumpulan data, proses pengolahan data dan perancangan aplikasi terdiri dari perancangan Aplikasi, basis model dan rancangan antar muka.

### **BAB IV      IMPLEMENTASI DAN ANALISIS HASIL PENGUJIAN**

Bab ini akan membahas mengenai implementasi dan pengujian dari rancangan sistem yang telah dirancang.

### **BAB V       KESIMPULAN DAN SARAN**

Bab ini berisi kesimpulan dan saran untuk pengembangan Tugas Akhir.