BAB 1
PENDAHULUAN

1.1 Latar Belakang

Short Message Service atau yang lebih dikenal dengan SMS merupakan sebuah fasilitas utama yang terdapat pada komunikasi *mobile* disamping fasilitas panggilan suara. Seiring dengan meningkatnya popularitas smart phone saat ini, SMS tidak hanya digunakan untuk saling berkomunikasi antara keluarga, sahabat, dan rekan kerja, melainkan dapat digunakan untuk tujuan yang lebih luas, seperti beriklan/berjualan bahkan dapat digunakan juga untuk hal yang negatif seperti penipuan. Beriklan dan melakukan penipuan melalui SMS adalah contoh dari SMS spam. SMS spam sendiri merupakan SMS yang tidak diharapkan dan tidak diinginkan yang masuk ke dalam perangkat *mobile* pengguna. SMS spam memiliki efek yang lebih besar kepada pengguna dibandingkan dengan email spam karena pengguna akan selalu melihat setiap SMS yang mereka terima, sehingga SMS spam dapat mempengaruhi pengguna secara langsung, terlebih lagi jika SMS spam yang mereka terima adalah SMS penipuan [10].

Untuk mengatasi masalah spam tersebut, salah satu cara yang dapat digunakan adalah dengan melakukan *filtering*. Pada dasarnya *filtering* ini akan membedakan antara SMS spam dan ham (bukan spam). Pada penelitian sebelumnya spam filtering lebih banyak diterapkan untuk pesan email dengan menggunakan beberapa algoritma klasifikasi teks seperti Naïve Bayes, Decision Tree, Support Vector Machine, dan Neural Network. Pada tugas akhir ini, penelitian tentang filtering spam akan difokuskan pada pesan SMS. SMS dan email memiliki beberapa perbedaan diantaranya ialah SMS hanya bisa mengirim dan menerima teks, sedangkan email bisa mengirim selain teks seperti gambar dan suara. SMS memiliki kapasitas teks yang terbatas jika dibandingkan dengan email. Selain itu pada SMS sering menggunakan singkatan-singkatan atau istilah-istilah sehingga dapat menyebabkan pengertian yang ambigu.

Metode yang akan digunakan dalam melakukan filtering SMS spam ini ialah Artificial Immune System (AIS). Metode ini pada dasarnya bekerja seperti system imun yang terdapat pada tubuh manusia yang bisa membedakan antara sel-sel di dalam tubuh dan sel-sel asing di luar tubuh yang berbahaya. Pada system filtering ini sel-sel di dalam tubuh dianalogikan sebagai ham dan sel-sel asing dari luar tubuh sebagai spam. Untuk mendapatkan akurasi yang lebih tinggi, metode AIS ini akan menambahkan algoritma *Tokenization With Vectors* yang akan diterapkan untuk dataset pada saat preprocessing. Metode dan algoritma ini diyakini akan dapat menghasilkan persentase akurasi yang baik dalam memisahkan SMS spam dan juga ham.

1.2 Perumusan Masalah

Berdasarkan latar belakang dari permasalahan di atas, perumusan masalah yang dapat dibahas pada Tugas Akhir ini ialah:

1. Bagaimana menerapkan metode *Artificial Immune System* (AIS) dan algoritma *Tokenization With Vectors* untuk SMS spam *filtering*?
2. Bagaimana perbandingan akurasi SMS spam *filtering* yang dihasilkan antara metode *Artificial Immune System* (AIS) yang menggabungkan algoritma *Tokenization With Vectors* dan tanpa menggabungkan algoritma *Tokenization With Vectors*?

1.3 Tujuan

Berdasarkan perumusan masalah diatas, maka tujuan dari Tugas Akhir ini ialah:

1. Dapat menerapkan metode *Artificial Immune System* (AIS) dan algoritma *Tokenization With Vectors* untuk SMS spam *filtering*.
2. Dapat mengetahui perbandingan akurasi SMS spam *filtering* yang dihasilkan antara metode *Artificial Immune System* (AIS) yang menggabungkan algoritma *Tokenization With Vectors* dan tanpa menggabungkan algoritma *Tokenization With Vectors*.

1.4 Batasan Masalah

Batasan masalah Tugas Akhir ini ialah:

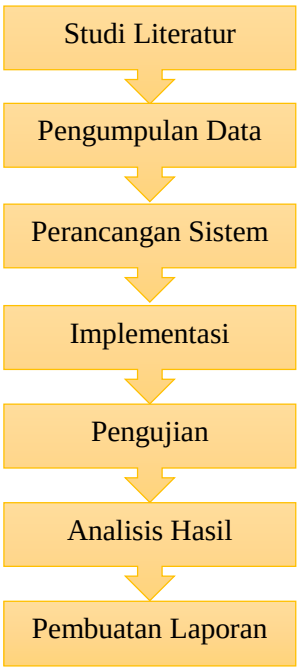
1. Dataset yang digunakan hanya yang berbahasa inggris
2. Dataset berbentuk *plain text* atau teks murni
3. Nomor pengirim tidak termasuk ke dalam proses klasifikasi
4. Program bersifat offline

1.5 Hipotesa

Adapun hipotesa dari Tugas Akhir ini ialah dengan menerapkan metode *Artificial Immune System* (AIS) dan algoritma *Tokenization With Vectors* pada SMS spam *filtering* akan dapat menghasilkan akurasi yang lebih baik dibandingkan tanpa menggunakan algoritma *Tokenization With Vectors* dalam membedakan antara SMS spam dan ham (bukan spam).

1.6 Metodologi Penyelesaian Masalah

Metodologi penyelesaian masalah yang akan digunakan untuk mendukung Tugas Akhir ini meliputi beberapa tahap, tahapan-tahapan tersebut terdapat pada gambar 1.1 di bawah ini:



Gambar 1.1 Tahapan Metodologi Penyelesaian Masalah

1.6.1 Studi Literatur

Pada tahap ini akan dilakukan proses untuk mendalami materi yang berhubungan dengan konsep atau teori dari SMS spam *filtering*, metode AIS, *tokenizing* dan beberapa teori pendukung lainnya. Studi literatur dapat dilakukan dengan mencari atau membaca dari berbagai sumber seperti perpustakaan dan internet.

1.6.2 Pengumpulan Data

Pada tahap ini akan dikumpulkan data berupa isi dari SMS berbahasa inggris yang mengandung spam dan ham. Data-data ini dapat diunduh secara gratis melalui internet. Data-data yang telah dikumpulkan ini nantinya akan diperlukan untuk proses *training* dan berguna untuk pengujian sistem.

1.6.3 Perancangan Sistem

Setelah memahami teori-teori dari permasalahan yang ada, maka tahap selanjutnya adalah melakukan perancangan sistem yang dapat menerima dan membedakan antara SMS spam dan ham. Sistem yang dirancang ini akan berguna untuk pengujian nantinya.

1.6.4 Implementasi

Pata tahap ini akan dibangun sebuah sistem simulasi berbasis Java yang telah menerapkan metode *Artificial Immune System* (AIS) dan algoritma *Tokenization With Vectors* berdasarkan perancangan yang telah dilakukan sebelumnya.

1.6.5 Pengujian

Pada tahap ini akan dilakukan pengujian terhadap sistem dengan menggunakan beberapa skenario. Cara pengujian yang akan dilakukan ialah dengan menggunakan data isi SMS yang telah dikumpulkan sebelumnya dan data-data tersebut akan diuji ke sistem sehingga sistem yang telah dibangun tersebut dapat membedakan antara SMS spam dan ham.

1.6.6 Analisis Hasil

Pada tahap ini akan dilakukan analisis terhadap pengujian yang telah dilakukan sebelumnya. Hasil yang dianalisis berupa persentase akurasi yang dihasilkan oleh sistem dalam menentukan suatu SMS tersebut termasuk spam atau ham.

1.6.7 Pembuatan Laporan

Tahap ini merupakan tahap akhir dimana tahap-tahap sebelumnya akan didokumentasikan dalam sebuah buku laporan.