

BAB I

Pendahuluan

1.1 Latar Belakang

Komunikasi menjadi hal yang sangat penting bagi kehidupan setiap orang, untuk saling berinteraksi satu dengan yang lain. Pada beberapa tahun terakhir, media pesan singkat atau yang lebih akrab disebut SMS (*Short Message Service*) menjadi media yang paling sering digunakan dalam komunikasi antar *mobile*[15]. Hal tersebut membuat sebagian orang yang tidak bertanggung jawab menggunakan fasilitas pesan singkat sebagai cara untuk menguntungkan dirinya sendiri, salah satunya yaitu digunakan sebagai media promosi berbagai produk atau jasa.

Bagi sebagian orang yang menganggap pesan singkat menjadi hal yang penting, mereka merasa terganggu akan adanya pesan-pesan singkat yang isinya hanya berupa promosi, layanan jasa atau pesan-pesan lain yang isinya tidak diinginkan. Pesan yang dikirimkan ke sembarang target secara langsung ataupun tidak langsung dengan isi yang tidak diharapkan oleh target ini disebut dengan *SMS Spam*[5]. Dari kondisi tersebut, tugas akhir ini membangun sebuah cara untuk melakukan *Spam filter*. *Spam Filter* merupakan sebuah teknik yang dibangun untuk secara otomatis mengidentifikasi Spam[5]. Dengan *SMS Spam filter* yang akan dibangun, maka setiap SMS yang masuk ke dalam sistem akan diidentifikasi dan diklasifikasikan.

Dalam proses identifikasi dan klasifikasi, yang perlu diperhatikan adalah ketika sebuah data telah dikumpulkan, proses klasifikasi belum bisa dilakukan secara langsung, ada tahap yang harus dilakukan yaitu tahap *preprocessing*. Tahap ini digunakan untuk membersihkan atau menyederhanakan data tanpa merubah informasi yang dikandungnya, sehingga dapat mengurangi waktu komputasi dalam proses klasifikasi[3]. Sifat SMS yang bebas dan tidak teratur karena bisa berupa kata-kata singkatan atau kata-kata tren anak muda, membuat tahap *preprocessing* menjadi wajib dan penting untuk dilakukan, agar dapat diperoleh data SMS yang lebih baik.

Dari sebuah penelitian tentang klasifikasi, menyebutkan algoritma *Multinomial Naïve Bayes* memiliki tingkat efektifitas tertinggi mencapai 98.22% dibandingkan dengan algoritma lain[7]. Dari hasil pengujian tersebut, maka dipilihlah algoritma *Multinomial Naïve Bayes* sebagai algoritma yang diterapkan dalam klasifikasi SMS pada *SMS Spam filter* di tugas akhir ini.

1.2 Rumusan Masalah

Dari kondisi yang ada dalam latar belakang, dirumuskan beberapa masalah yang akan dikaji pada tugas akhir ini:

1. Bagaimana *preprocessing* yang baik, untuk menangani isi SMS yang bersifat bebas dan tidak teratur?
2. Bagaimana menerapkan algoritma *Multinomial Naïve Bayes* pada sistem *SMS Spam filter*?
3. Bagaimana *SMS Spam Filter* mengidentifikasi *SMS spam* dan *not spam*?
4. Bagaimana mengukur tingkat akurasi dari *SMS Spam Filter*?

1.3 Tujuan

Tujuan yang akan dicapai dari tugas akhir ini adalah:

1. Dapat mengatasi permasalahan tentang isi SMS yang bersifat bebas dan tidak teratur.
2. Mengimplementasikan algoritma *Multinomial Naïve Bayes* pada SMS Spam Filter.
3. Mengidentifikasi *SMS Spam* dan *not spam*.
4. Menganalisa tingkat akurasi algoritma *Multinomial Naïve Bayes* pada *SMS Spam Filter*.

1.4 Batasan Masalah

Batasan masalah yang akan dibahas dalam tugas akhir ini:

1. Jenis Preprocessing data yang akan digunakan adalah : *Slang Handling*, *Tokenization*, *Stopwords elimination* dan *Stemming*.
2. Untuk *Slang Handling* menggunakan data *dictionary slang*[12].
3. Untuk *Stopwords elimination* menggunakan 2 *stopword list*:
 - Short (default English stopword list) [10].
 - Long (long stopword list) [10].
4. Klasifikasi SMS dibagi menjadi dua kategori yaitu *spam* dan *ham (not spam)*.
5. SMS yang akan dianalisa adalah SMS berbahasa Inggris.
6. Dataset yang digunakan berasal dari 2 sumber :
 - The SMS Spam Collection v.1[13].
 - British SMS[9].

dan akan dikomposisikan sesuai keperluan pengujian, dalam proses training maupun dalam proses testing.

1.5 Metodologi

Metodologi yang digunakan dalam penyelesaian masalah :

1. Studi Literatur
Merupakan tahap pemahaman materi, identifikasi masalah, serta pencarian solusi yang berkaitan dengan masalah tersebut.
2. Pengumpulan dan Pengolahan Data
Merupakan tahap pengumpulan data-data yang berupa SMS, baik spam ataupun tidak spam. Yang kemudian dilakukan beberapa pengolahan untuk digunakan sebagai data training dan data testing.
3. Perancangan dan Implementasi Sistem
Merupakan tahap menganalisa kebutuhan dan keperluan sistem agar sesuai dengan permasalahan yang akan diselesaikan. Yang kemudian dilanjutkan dengan implementasi dari perancangan yang dibuat berdasarkan analisa kebutuhan sistem.
4. Pengujian Sistem
Merupakan tahap pengujian dari sistem, dengan menggunakan data-data yang telah didapatkan.
5. Analisis Hasil dan Kesimpulan
Menganalisa hasil dari pengujian sistem yang telah dilakukan, berdasarkan skema uji yang sudah dibuat. Kemudian merumuskan kesimpulan dari hasil analisa tersebut.
6. Dokumentasi
Merupakan tahap yang dilakukan dari awal sampai akhir penelitian. Dengan mencatat hal-hal yang penting dalam setiap pelaksanaan penelitian.

1.6 Sistematika Penulisan

Sistematika penulisan dalam Tugas Akhir ini adalah sebagai berikut :

1. BAB I Pendahuluan
Pada BAB I diuraikan isi dan rencana dari Tugas Akhir secara menyeluruh, yang meliputi beberapa sub bab : Latar Belakang, Perumusan Masalah, Tujuan, Batasan Masalah, Metodologi dan Sistematika Penulisan.
2. BAB II Tinjauan Pustaka
Pada BAB II berisi tentang materi-materi pendukung yang digunakan sebagai landasan teori untuk membangun kinerja setiap proses dalam sistem.
3. BAB III Perancangan dan Implementasi
Pada BAB III berisi tentang perancangan dan penjelasan umum sistem secara keseluruhan, dan juga proses-proses yang dilakukan dalam

setiap bagiannya. Selain itu, proses-proses juga dijelaskan secara detail dengan pemberian contoh kasus yang berupa input dan outputnya.

4. BAB IV Pengujian dan Analisis

Pada BAB IV berisi tentang tujuan dari pengujian, bentuk pembagian data yang digunakan dalam pengujian, penjelasan skenario pengujian, hasil dari pengujian serta analisis dari hasil yang telah didapatkan dalam pengujian.

5. BAB V Penutup

Pada BAB V ini berisi kesimpulan serta saran yang telah didapatkan dari hasil Tugas Akhir secara keseluruhan.