

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Teknologi komunikasi digunakan untuk dapat saling berinteraksi satu dengan yang lain. Telepon selular merupakan contoh teknologi komunikasi yang banyak digunakan oleh masyarakat diseluruh dunia dalam beberapa dekade belakangan ini dan di Indonesia pengguna telepon seluler sekitar 152 juta pengguna pada tahun 2009 [7]. Kepopuleran telepon seluler membawa salah satu layanannya tersebut naik ke level yang lebih tinggi yaitu menjamurnya penggunaan pesan singkat atau yang lebih dikenal dengan SMS (*Short Message Service*). Kelebihan SMS yang ditawarkan yaitu merupakan layanan telepon seluler yang berfokus pada pertukaran pesan tulisan yang murah [4].

Pengguna yang banyak dan kelebihan yang ditawarkan oleh layanan SMS dimanfaatkan oleh beberapa pihak untuk mendapatkan keuntungan baik secara personal maupun setingkat industri, seperti pengiriman pesan iklan komersial, penipuan pengisian pulsa, pemenang undian atau transaksi keuangan palsu kepada pengguna layanan SMS [17]. SMS tersebut dapat dikategorikan sebagai SMS sampah (*SMS Junk* atau *Spam*) yang mengakibatkan terganggunya pengguna telepon seluler seperti selalu munculnya notifikasi SMS *spam* yang dianggap seharusnya tidak diperlukan dan juga berefek terhadap kepadatan trafik operator seluler [6].

Pengklasifikasian SMS merupakan langkah yang dilakukan untuk menyaring SMS menurut kelasnya. *Collaborative Naïve Bayes* dan *Content-based Naïve Bayes* digunakan sebagai metode untuk mengklasifikasikan SMS. *Collaborative Naïve Bayes* merupakan teknik pengklasifikasian berdasarkan opini beberapa *user* untuk memprediksi kelas SMS dari seorang *user*. [10]. Ketika SMS tidak mempunyai kelas dan belum ada user yang merekomendasikan kelas SMS tersebut, maka *Content-Based Naïve Bayes* digunakan untuk membantu mengklasifikasikan SMS dan menghindari *sparse* yang terlalu besar. SMS tidak memiliki aturan umum, sehingga SMS memiliki beberapa karakteristik seperti SMS yang tidak teratur, tidak ada singkatan, dan SMS formal, oleh karena itu *preprocessing* digunakan untuk membersihkan dan membuat SMS menjadi lebih teratur sehingga informasi dari SMS dapat digunakan untuk mengklasifikasikan SMS. *Preprocessing* yang digunakan antara lain penanganan singkatan (*slang replacing*), merubah kata menjadi kata dasar (*stemming*), dan menghapus kata yang tidak informatif (*feature extraction* dan *selection*).

Dari sebuah penelitian sebelumnya, didapatkan akurasi dari *Collaborative Naïve Bayes* sebesar 68.5% untuk 50 user [5] dan 98.7% pada *Content-Based Naïve Bayes*[4].

## 1.2 Rumusan Masalah

Dari paparan latar belakang yang telah dideskripsikan di atas, maka dapat dirumuskan beberapa permasalahan yang menjadi objek penelitian pada Tugas Akhir ini, yaitu:

1. Bagaimana menerapkan metode *Collaborative Naïve Bayes* dan *Content-Based Naïve Bayes* untuk mengklasifikasikan SMS?
2. Bagaimana menerapkan *preprocessing* yang baik untuk mendapatkan informasi yang berguna pada SMS?
3. Bagaimana mengukur tingkat keakurasian pengklasifikasian SMS?

## 1.3 Tujuan

Adapun tujuan yang diharapkan dari dilakukannya penelitian Tugas Akhir ini dari pemaparan latar belakang serta perumusan masalah yang didapat adalah sebagai berikut:

1. Dapat menerapkan metode *Collaborative Naïve Bayes* dan *Content-Based Naïve Bayes* untuk mengklasifikasikan SMS.
2. Dapat mengimplementasikan *preprocessing* yang baik sehingga informasi yang dibutuhkan untuk klasifikasi didapatkan.
3. Dapat mengukur dan menganalisis tingkat akurasi pengklasifikasian SMS.

## 1.4 Batasan Masalah

Batasan masalah yang dibahas dalam penelitian tugas akhir ini:

1. SMS yang akan diklasifikasikan adalah berbahasa Inggris
2. Bahasa *user* yang memberikan kelas untuk *Collaborative Naïve Bayes* pada SMS berbahasa Inggris adalah bahasa Indonesia.
3. Jumlah *user* yang memberikan rekomendasi kelas sebanyak 40 *user*.
4. Klasifikasi SMS dibagi menjadi dua kategori yaitu *spam* dan *ham*.
5. Terdapat tiga tahapan *Preprocessing* yang digunakan: *Slang Handling*, *Stopword Removal*, dan *Stemming*.
6. *Database* yang digunakan menggunakan MySQL dengan bantuan aplikasi *PhpMyAdmin* dengan server lokal yaitu XAMPP.
7. Untuk *Stopwords Removal* menggunakan kamus *stopwords* [9].
8. Untuk *Slang Handling* menggunakan kamus *slang* berbahasa Inggris [12].
9. *Dataset* yang digunakan adalah *SMSSpamCollection* yang diperoleh dari UCI Machine Learning Repository [16].
10. Rekomendasi kelas didapatkan dengan menyebarkan survey ke beberapa responden.

## 1.5 Metodologi

Untuk menyelesaikan permasalahan dan mencapai tujuan yang diharapkan, maka terdapat beberapa metodologi yang diterapkan sebagai pemecah permasalahan, yaitu sebagai berikut:

1. Studi Literatur

Tahap ini berupaya untuk mendapatkan serta mengumpulkan referensi permasalahan serta pengidentifikasian terhadap masalah yang masih mungkin terjadi yang akan menjadi solusi yang dapat diterapkan pada penelitian Tugas Akhir ini. Penulis mencari pengetahuan tentang *filtering text*, metodologi yang sesuai dari penelitian-penelitian sebelumnya, *Collaborative* dan *Content-Based filtering* dan pemanfaatannya. Sumber yang digunakan adalah jurnal, buku, serta situs internet dengan penulis yang sah.

2. Pengumpulan Data dan Analisis Data Awal

Penulis mengumpulkan data-data sebagai pendukung penelitian Tugas Akhir ini. Data disebar sebagai survey kepada beberapa responden untuk mendapatkan kelas rekomendasi. Pada data *Content-Based* akan dibagi menjadi data latih dan data uji yang masing-masing memiliki SMS *ham* dan *spam* di dalamnya. *Pre-processing* data (SMS) dan *extraction of features* dilakukan.

3. Perancangan Sistem

Pada tahap ini, penulis akan menentukan bagaimana rancangan sistem yang diperlukan sehingga kebutuhan sistem untuk dapat menganalisis data dan menyelesaikan permasalahan dapat diimplementasikan. Sistem dirancang sesuai dengan skema *Collaborative Filtering* dan *Content-Based* sebagai pendukungnya dan meningkatkan akurasi.

4. Implementasi Sistem

Sistem dibangun dengan menggunakan metode *Naïve Bayes*.

5. Pengujian Sistem

Skema pengujian yang dilakukan dengan melihat keakurasian sistem dalam mengeluarkan hasil yang diinginkan antara SMS *spam* dengan SMS *ham*. Pengujian difokuskan terhadap jumlah *user* yang memberikan rekomendasi.

6. Analisis Hasil

Tahap ini akan melihat keakurasian keluaran yang diberikan sebagai hasil yang diproses oleh sistem dan menjadi kesimpulan terhadap data masukan (SMS).

7. Penulisan Laporan

Melakukan penulisan terhadap penelitian mulai dari sebagian tahap pengumpulan data hingga analisis hasil uji masukan yang diberikan kepada sistem yang menjadi kesimpulan Tugas Akhir ini.

## **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.
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 diperoleh.
5. **BAB V Penutup**  
Pada BAB V berisikan kesimpulan serta saran yang telah didapatkan dari hasil Tugas Akhir secara keseluruhan.