

1. Pendahuluan

1.1 Latar Belakang

Sekarang ini, SMS merupakan media komunikasi yang paling sering digunakan. Seiring dengan meningkatnya intensitas penggunaan SMS ini dimanfaatkan oleh sebagian orang yang tidak bertanggung jawab untuk melakukan tindak kriminal seperti penipuan melalui SMS. Maraknya aksi penipuan ini mengakibatkan keresahan dan ketidaknyamanan bagi penerima SMS. SMS tersebut dikenal dengan SMS spam. Salah satu SMS yang berindikasi spam adalah SMS yang berisi permintaan pulsa telepon seluler ke nomor tertentu dengan mengatasnamakan mama atau papa. SMS penipuan yang beredar luas dimasyarakat cenderung memiliki pola tertentu. Hanya saja masyarakat awam tidak banyak mengetahui sehingga tertipu oleh SMS tersebut. Oleh karena itu tugas akhir ini membangun sebuah cara untuk melakukan SMS spam *filter* yang tidak diinginkan penerima. SMS spam *filter* merupakan teknik yang dibangun untuk mengklasifikasikan SMS spam atau bukan spam (ham) secara otomatis. Dengan mengenali pola SMS dari SMS spam *filter* yang akan dibangun, maka setiap SMS yang masuk ke dalam sistem akan diidentifikasi dan diklasifikasikan, dan masyarakat dapat lebih berhati-hati dalam menindaklanjuti SMS yang diterima sehingga kejahatan melalui SMS ini dapat dihindari.

Dalam proses identifikasi dan klasifikasi, data yang telah dikumpulkan perlu melakukan *preprocessing* terlebih dahulu, guna membersihkan atau menyederhanakan data tanpa merubah informasi yang dikandungnya, sehingga dapat mengurangi waktu komputasi dalam proses klasifikasi[SAN07]. Karakteristik SMS yang bersifat bebas dan tidak teratur karena bisa berupa kata-kata singkatan atau kata-kata trend anak muda, membuat tahap *preprocessing* menjadi penting untuk dilakukan, agar dapat diperoleh data SMS yang lebih terstruktur.

Penelitian sebelumnya tentang klasifikasi sudah banyak dilakukan dengan berbagai teknik seperti SVM, *Naïve Bayes*, ID3, C45. Akan tetapi tingkat efektivitas dari setiap teknik berbeda-beda. Pada penelitian tersebut menyebutkan algoritma *Multinomial Naïve Bayes* memiliki tingkat efektivitas tertinggi mencapai 98.2% dibandingkan dengan metode lain[MAT11]. 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 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, terdapat beberapa permasalahan yang muncul pada tugas akhir ini, yaitu sebagai berikut:

- a. Bagaimana *preprocessing* yang baik, untuk menangani isi SMS yang bersifat bebas dan tidak teratur?
- b. Bagaimana klasifikasi SMS spam *filter* terhadap identifikasi SMS?
- c. Bagaimana implementasi pada SMS spam *filter* berbahasa Indonesia dengan menggunakan algoritma *Multinomial Naïve Bayes* pada sistem *smartphone* Android?
- d. Bagaimana mengukur kinerja dari SMS spam *filter*?

1.3 Batasan Masalah

Adapun beberapa batasan masalah yang muncul pada tugas akhir ini, yaitu sebagai berikut:

- a. Sistem *filter* SMS spam hanya menggunakan bahasa Indonesia.
- b. Data SMS bernilai spam dan bukan spam (ham) yang digunakan untuk proses *training* berasal dari hasil penelitian sebelumnya (Karimul Makhtidi, 2012), ditambah dengan data primer yang dikumpulkan penulis sehingga total data menjadi 180 data SMS.
- c. Sistem berbasis Android.
- d. Parameter yang menjadi pembanding yaitu *accuracy*, *precision*, *recall*.

1.4 Tujuan

Tujuan dari tugas akhir ini untuk:

- a. Dapat mengatasi permasalahan tentang isi SMS yang bersifat bebas dan tidak teratur.
- b. Dapat mengklasifikasi SMS spam *filter* terhadap identifikasi SMS.
- c. Dapat mengimplementasi algoritma *Multinomial Naïve Bayes* pada sistem SMS spam *filter* berbahasa Indonesia pada *smartphone* Android.

- d. Mengukur kinerja sistem SMS spam *filter* berbasis algoritma *Multinomial Naïve Bayes*.

1.5. Metodologi Penyelesaian Masalah

Pembuatan tugas akhir ini dilakukan dengan menggunakan metodologi penyelesaian masalah sebagai berikut:

1. Studi Literatur

Studi literatur adalah proses dimana akan dilakukan pencarian, pengumpulan, penyaringan, pembelajaran, pemahaman dan pendalaman literatur yang berhubungan dengan SMS, *filtering*, algoritma *Multinomial Naïve Bayes*, TF-IDF, spam. Beberapa literatur berupa buku dan *paper* yang mendukung teori dan konsep tersebut seperti berikut:

- a. *Hello, Android, Introducing Google's Mobile Development Platform*
 - b. *Content based Hybrid SMS Spam Filtering System*
 - c. *Data Mining The Web*
 - d. Klasifikasi Teks Pesan Spam Menggunakan Algoritma *Naïve Bayes*.
 - e. *Data Mining: Concept and Techniques*
 - f. *Introduction to Information Retrieval*
 - g. *A Comparison of Event Models for Naive Bayes Text Classification*
 - h. *Spam Filtering with Naive Bayes – Which Naive Bayes?*
 - i. *Text Mining Classifier Untuk Program Studi Tesis Text Mining Dengan Metode Naïve Bayes Classifier dan Support Vector Machines Untuk Sentiment Analysis*
 - j. ***Android Application Development: Programming with the Google SDK***
 - k. *SMS Spam Filtering: Methods and Data*
 - l. *Simple SMS Spam Filtering on Independent Mobile Phone*
- ### 2. Tahap Pengumpulan Data
- Pengumpulan data dilakukan secara manual dari *mobile phone end user* kemudian akan dimasukkan ke dalam database untuk dianalisis.

3. *Preprocessing* Data

Data yang sudah terkumpul kemudian dilakukan tahapan *tokenizing*, *slang handling*, *stopwords*, dan *stemming* kemudian selanjutnya diproses kedalam klasifikasi SMS spam *filter*

4. Implementasi

Implementasi dari tugas akhir ini adalah pembobotan seleksi fitur. Pembobotan seleksi fitur yang dilakukan yaitu pembobotan TF-IDF. Kemudian diuji untuk diklasifikasi dengan menggunakan algoritma *Multinomial Naïve Bayes*. Dengan demikian sistem akan mampu mem-*filter* SMS kemudian mengklasifikasi SMS spam *filter* terhadap identifikasi SMS.

5. Analisis Hasil dan Kesimpulan

Analisis hasil pengujian dan kesimpulan dari tugas akhir ini adalah menganalisis hasil pengujian sistem yang dapat mem-*filter* SMS tersebut spam atau bukan spam (ham).

6. Pembuatan Laporan

Pembuatan laporan dari tugas akhir ini akan dilakukan berupa dokumentasi dari hasil pengujian sistem tugas akhir sesuai dengan penulisan yang benar dan sesuai ketentuan - ketentuan atau sistematika yang telah ditetapkan.