

1. Pendahuluan

1.1 Latar Belakang

Pesan singkat atau *short message service* (SMS) adalah salah satu layanan komunikasi untuk mengirim dan menerima pesan singkat berupa teks pada telepon seluler (ponsel). SMS masih digunakan setiap harinya karena kemudahan penggunaan, sederhana, cepat, dan murah. Saat ini selain digunakan untuk komunikasi teks antarpengguna ponsel, SMS digunakan untuk berbagai kebutuhan seperti otentifikasi *mobile banking*, dan sinkronisasi media sosial (Facebook, Twitter). Meningkatnya penggunaan SMS dimanfaatkan oleh banyak pihak untuk mendapatkan keuntungan, salah satunya adalah mengirimkan *spam* melalui SMS. SMS spam yang beredar luas dimasyarakat cenderung memiliki pola tertentu. Hanya saja masyarakat awam tidak banyak mengetahui sehingga tertipu oleh SMS tersebut.

SMS *Spam*, didefinisikan oleh *The Spam Track Text Retrieval Conference* (TREC) adalah suatu konten yang tidak diminta dan tidak dikehendaki yang dikirim sembarang, baik langsung maupun tidak langsung oleh pengirim yang tidak memiliki hubungan dengan penerima [1]. SMS *Spam* dibagi menjadi 3 jenis yaitu, iklan, promosi dan penipuan. Bagi kebanyakan orang hal ini sangatlah mengganggu sehingga membuat penanganan terhadap SMS *spam* sangat penting untuk dilakukan. Minimnya koleksi atau *sample* data SMS *spam* bahasa Indonesia, mengakibatkan SMS *spam* filtering sulit dikembangkan di Indonesia. Dari kondisi tersebut, maka pada tugas akhir ini, dibuatlah sebuah *corpus* SMS *spam* bahasa Indonesia untuk studi SMS *spam filtering* yang akan sekaligus diuji performansinya untuk sistem *learning SMS spam filtering*.

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut, maka dapat dirumuskan permasalahan sebagai berikut.

- a. Bagaimana cara mendapatkan data SMS *spam*?
- b. Bagaimana menguji performansi *corpus* SMS *spam* bahasa Indonesia?
- c. Bagaimana mengukur hasil evaluasi pada *corpus* SMS *spam* bahasa Indonesia?

1.3 Tujuan

Tujuan dari pembuatan tugas akhir ini adalah :

- a. menghasilkan *corpus* SMS *spam* bahasa Indonesia,
- b. Pengujian *corpus* SMS *spam* bahasa Indonesia dengan metode klasifikasi *Decision Tree* dan *Naïve Bayes*,
- c. menjelaskan cara mengukur hasil evaluasi *corpus* SMS *spam* bahasa Indonesia.

1.4 Batasan Masalah

Data SMS yang dikumpulkan dalam pengembangan *corpus* hanya memiliki konten bahasa Indonesia. Karena dalam studi literatur yang dilakukan penulis, di Indonesia belum tersedia *corpus* SMS *spam* yang memiliki konten bahasa Indonesia.

Penentuan *atribut* dalam data SMS dilakukan manual yang akan digunakan dalam menghitung *entropy* dan menghasilkan *model tree*.

Corpus SMS *spam* diuji dengan metode klasifikasi, yaitu *Decision Tree* berbasis *web*, dan dengan *tool RapidMiner* menggunakan metode *Naïve Bayes*.

1.5 Metodologi Penyelesaian Masalah

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

a. Identifikasi masalah

Tahap awal yaitu dengan melakukan identifikasi terhadap permasalahan yang ada.

b. Studi literatur

Studi literatur meliputi pencarian materi dengan mempelajari referensi, baik berupa makalah, jurnal, maupun buku yang relevan dan tentunya membahas tentang *crowdsourcing*, *data mining* dan *corpus*.

c. Tahap pengumpulan data

Tahap ini meliputi pembukaan akses input SMS menggunakan *tool* SMS gateway *ProgramSMS*. Pengguna ponsel yang menerima SMS *spam* dapat berpartisipasi dengan mengirim SMS *spam* tersebut ke nomor SMS yang digunakan pada SMS gateway penulis. Data SMS yang dikumpulkan dilakukan pelabelan sesuai kelasnya, yaitu *spam* dan *ham*, sehingga menghasilkan sebuah *corpus* SMS *spam*. Pengujian *corpus* akan dilakukan dengan membagi data SMS menjadi data *training* dan *testing*.

d. Implementasi metode

Pada tahap ini *corpus* SMS *spam* akan diuji dengan beberapa model klasifikasi, yaitu *Decision Tree* berbasis *web program*, dan dengan *tool* *RapidMiner* menggunakan metode *Naïve Bayes*.. Pengujian akan menghasilkan 2 kelas prediksi, yaitu *spam* dan *ham* untuk pengujian *corpus*.

e. Kesimpulan

Kesimpulan dari tugas akhir ini adalah mengevaluasi *corpus* dengan metode *Decision Tree* yang akan sekaligus diuji performansinya dan dengan *tool* *RapidMiner* menggunakan metode *Naïve Bayes*. Parameter pengukuran hasil evaluasi yang digunakan, yaitu *accuracy*, *precision*, *recall*.

f. 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.