

Abstrak

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. 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. 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*.

Data SMS dikumpulkan dengan model *crowdsourcing* yang melibatkan pengguna ponsel. Data SMS yang terkumpul dilakukan pendekatan linguistik yang divalidasi oleh ahli Bahasa Indonesia untuk menentukan kelas *spam* dan *ham*, sehingga menghasilkan sebuah *corpus* SMS *spam*. *Corpus* SMS *spam* akan diuji performansinya dengan metode klasifikasi *Decision Tree* berbasis *web program* dan dengan *tool RapidMiner* menggunakan metode *Naïve Bayes*. Pengujian tersebut menghasilkan kesimpulan bahwa data SMS pada *corpus* memiliki karakter yang kuat sesuai kelasnya masing-masing, dengan rata-rata *accuracy* 92,91% dari kedua pengujian.

Kata kunci: SMS *spam*, SMS *ham*, *corpus*, Bahasa Indonesia, *crowdsourcing*, *Decision Tree*, *Naïve Bayes*, *RapidMiner*.