

Abstrak

Maraknya penggunaan SMS membuat pihak tertentu memanfaatkan hal tersebut dengan menyebarkan SMS ke berbagai pihak demi keuntungannya sendiri. SMS tersebut disebut juga dengan SMS Spam. Bagi sebagian orang, hal tersebut sangat mengganggu. Oleh sebab itu, maka dibangunlah SMS Spam Filter yang bertujuan untuk menyaring SMS. Dalam tugas akhir ini, dilakukan penelitian tentang SMS Spam Filter, yang akan mengklasifikasikan SMS menjadi Spam dan Ham (Tidak Spam). Isi SMS yang berupa teks dan cenderung tidak teratur, membuat data yang akan digunakan untuk klasifikasi perlu dilakukan preprocessing terlebih dahulu. Preprocessing yang digunakan dalam tugas akhir ini adalah Slang Handling, Stopword Elimination, Stemming, dan Tokenization. Setelah semua data dilakukan preprocessing maka algoritma yang akan mengklasifikasikan SMS-nya adalah algoritma multinomial naïve bayes. Tugas akhir ini akan menggunakan 2 sumber data yaitu: The SMS Spam Collection v.1 dan British SMS. Yang keduanya akan dikomposisikan sesuai pengujian. Pengujian yang dilakukan yaitu membandingkan beberapa skenario pengujian berdasarkan penggunaan preprocessingnya. Dan telah didapatkan hasil terbaik dengan akurasi mencapai 98.15% dengan pemilihan preprocessing Stopword Elimination dengan Stemming.

Kata Kunci : Klasifikasi, SMS Spam Filter, Slang, Stopword Elimination, Stemming, Multinomial Naïve Bayes.