

Abstrak

Masalah utama dari sebuah forum adalah adanya pengguna yang sering mengomentari suatu thread atau topik tetapi apa yang dikomentari tidak berguna atau tidak berhubungan sama sekali dengan isi thread atau topik tersebut. Komentar atau postingan seperti ini dikenal dengan *junk post*. Banyak sekali kerugian akibat adanya *junk post*, diantaranya adalah menambah halaman topik tersebut yang mengakibatkan user lain sulit untuk membaca seluruh topik serta menambah waktu membaca forum, mengubah arah pembicaraan sehingga bisa berkembang *junk post* lainnya dari user lain, pokok topik yang penting bisa dikaburkan oleh *junk post* sehingga ada beberapa hal yang terlewat, dan berbagai macam masalah lainnya.

Untuk itulah penelitian ini bertujuan untuk mengklasifikasikan suatu postingan atau komentar termasuk *junk* atau tidak, sehingga diharapkan kedepannya dapat mengurangi *junk post* pada forum internet. Metode yang akan digunakan untuk penelitian ini adalah dengan menggunakan metode ***K-Nearest Neighbor untuk pengklasifikasian teks***. Dalam penerapannya, dilakukan juga pembobotan postingan dengan *tf-idf* dan untuk menghitung nilai *similarity* dari postingan yang ada dengan thread atau topik yang berkaitan menggunakan *Pearson Correlation Distance*.

Hasil dari penelitian ini menunjukkan bahwa nilai *f1-measure* yang didapat untuk skenario pengujian 1 adalah 70.71% dengan jumlah $k=93$, untuk skenario pengujian 2 adalah 70.00% dengan jumlah $k=61$, dan untuk skenario pengujian 3 adalah 71.72% dengan jumlah $k=41$. Pada masing-masing skenario, *f1-measure* yang lebih baik dihasilkan ketika tidak menggunakan *stop word removal*. Hal ini di karenakan postingan dalam forum bersifat tidak baku dan memiliki kata yang sangat variatif sehingga kurang cocok apabila *stop word removal* digunakan pada penelitian ini.

Kata kunci: *text classification, junk post, k-nearest neighbor, pearson correlation distance, tf-idf.*