

Abstrak

Seiring dengan penggunaan internet sebagai media untuk bertukar informasi, bertukar dan berbagi opini juga dapat dilakukan. Dari opini-opini yang ada di internet, tidak semua opini tersebut berguna sehingga terkadang menyulitkan pihak-pihak yang membutuhkan informasi yang terkait. Opini yang tidak diperlukan tersebut dapat disebut juga dengan *opinion spam*. *Spam opinion* dalam dunia internet dapat terjadi dalam hal apapun, salah satunya dalam forum. *Spam opinion* dalam forum dapat berupa opini-opini sampah yang tidak berguna dalam suatu topik permasalahan yang sedang dibahas.

Text Classification merupakan salah satu solusi yang dapat mengatasi permasalahan tersebut. Salah satu metode pada *Text Classification* adalah *Support Vector Machine*. Klasifikasi berfokus pada penentuan postingan yang termasuk *junk post* atau tidak. Dalam Tugas Akhir ini, proses pembobotan dataset dengan memanfaatkan nilai TF dan juga nilai TF-IDF. Pembobotan dengan TF-IDF menghasilkan hasil yang lebih baik jika dibandingkan dengan TF saja. Kedua pembobotan tersebut digunakan untuk menghitung nilai *similarity* dari masing-masing postingan dan dikaitkan dengan topik yang sedang dibahas pada forum dengan menggunakan *cosine similarity*. Hasil evaluasi menunjukkan metode *Support Vector Machine* mampu menyelesaikan permasalahan spam opinion ini. Proses preprocessing data seperti penghilangan *stopword* dapat menurunkan hasil akurasi klasifikasi, sedangkan nilai parameter C tidak dapat mempengaruhi hasil klasifikasi secara signifikan.

Kata kunci: *text classification, opinion spam, Support Vector Machine, TF, TF-IDF, cosine similarity*