

Abstrak

Banyak algoritma yang dapat digunakan untuk mengatasi masalah *email spam*, contohnya *Naïve Bayes*. Namun, dengan hanya menggunakan *classifier* tunggal maka *filtering system* dapat dengan mudah diserang oleh pembuat *spam*. Oleh karena itu, diterapkan algoritma *Symbiotic Naïve Bayes* yang menggunakan *local filter* dari *user* yang berbeda untuk meningkatkan performansi *filtering* pada level *personal*.

Dilakukan beberapa pengujian untuk melihat pengaruh mekanisme pembobotan kata, *ratio data training* dan *data testing*, dan jumlah *dataset* terhadap performansi *filtering system* yang diukur dengan nilai *precision*, *recall*, dan *f - measure*. Ketiga hal tersebut dilakukan untuk mengetahui mekanisme pembobotan yang lebih baik, pada kondisi *ratio data training* dan *data testing* yang seperti apa, dan berapa jumlah ideal dari *dataset* sehingga *Symbiotic Naïve Bayes* dapat menghasilkan performansi yang tinggi.

Berdasarkan hasil pengujian dapat disimpulkan bahwa *filtering system* yang menggunakan data hasil mekanisme pembobotan *term transformation* menghasilkan performansi yang lebih baik dibandingkan dengan *term frequency*. Pada saat *ratio data training* lebih besar dibandingkan dengan *data testing* maka *filtering system* akan menghasilkan performansi yang lebih baik. Dan penambahan jumlah *dataset* dapat meningkatkan juga performansi *filtering system*.

Kata Kunci: *Naïve Bayes*, *Symbiotic Naïve Bayes*, Klasifikasi, *Dataset*, *Data Training*, *Data Testing*, *Term Frequency*, *Term Transformation*, *Ratio Data Training* dan *Data Testing*, *Precision*, *Recall*, *F – measure*