

Abstrak

Orang menghabiskan banyak waktu untuk membaca *email* dan memutuskan apakah *email* itu *spam* atau *non-spam*. Beberapa orang lainnya menghabiskan beberapa tambahan waktu untuk memberi label pada *email* mereka yang akan digunakan untuk *men-training local spam filters* yang ada di masing-masing komputer mereka. Namun di sisi lain *email service provider* mencoba untuk meringankan mereka dengan menggunakan *spam filters* pada server mereka. Dimana pada *spam filters* yang bersifat *server-based* ini tidak dapat menggunakan *labeled email* dari individual user, melainkan sumber yang ada secara umum, seperti *email newsgroup* atau *email-email* yang ditandai melalui *spam traps*. Sedangkan dari tiap individu pengguna *email* tersebut memiliki karakteristik yang berbeda dalam menyeleksi *email* yang dianggap *spam* ataupun *non-spam*. Di dalam tugas akhir ini, *data mining* digunakan untuk memutuskan apakah suatu *email* yang diterima oleh user/pengguna *email* adalah suatu *email spam* atau *non-spam* dan menggunakan ECML PKDD 2006 Discovery Challenge Data Mining Competition sebagai studi kasus dengan data *email* yang telah dikodekan menjadi *bag-of-word vector space* sehingga tidak diketahui secara pasti bentuk sebenarnya dari *email* tersebut. Dan metoda *Evolving Fuzzy Classifier* digunakan untuk klasifikasi menentukan jenis dari *email* dengan algoritma genetika sebagai metoda pengembangan atau evolusi bagi *Fuzzy Classifier* yang ada.

Kata kunci : *data mining, klasifikasi ,evolving fuzzy classifier , spam, genetic algorithm, non-spam, email, bag-of-word vector space, fuzzy classifier.*