

ABSTRAK

Berkembangnya kemajuan teknologi komunikasi saat ini memudahkan orang-orang untuk berkomunikasi. Salah satu teknologi komunikasi yang sering digunakan adalah email. Email sendiri adalah suatu metode pertukaran pesan-pesan digital antara satu orang pembuat pesan ke satu atau banyak penerima. Email sendiri sekarang sering disalahgunakan untuk kepentingan pribadi, kelompok atau organisasi. Terkadang seseorang pemilik email menerima pesan yang tidak diinginkan dari orang yang tidak dikenal. Email ini disebut sebagai email spam. Salah satu cara untuk menangani ini adalah melakukan pemfilteran email spam berdasarkan konten. Metode learning pohon keputusan dengan algoritma C4.5 dapat digunakan untuk menangani permasalahan tersebut. Untuk mendapatkan kinerja pohon keputusan yang baik dapat menggunakan metode ensemble Random Forest. Random Forest menggunakan kombinasi dari beberapa pohon keputusan dengan inputan jumlah feature yang akan digunakan pada pohon keputusan. Penggunaan jumlah feature yang tepat akan menghasilkan nilai akurasi yang baik dalam memfilter data spam atau ham.

Kata kunci : *pemfilteran email, pohon keputusan, algoritma C4.5, metode ensemble, random forest, akurasi.*