

Abstrak

Perkembangan teknologi internet sangat cepat membuat jumlah informasi berupa dokumen teks semakin banyak, oleh karena itu diperlukan suatu metode yang memudahkan pembaca untuk mencari informasi melalui proses kategorisasi. Namun tingginya dimensi data dapat mengganggu performansi hasil kategorisasi. Oleh karena itu dibutuhkan pemilihan *feature* yang berpengaruh besar terhadap kategorisasi yaitu *feature selection*. Ada beberapa algoritma dalam *feature selection* salah satunya yaitu *Ambiguity Measure*. Pada tugas akhir ini mengimplementasikan algoritma *feature selection* yaitu AM (*Ambiguity Measure*), dilakukan analisis hasil *feature selection* dengan menggunakan nilai *threshold* untuk memilih *feature-feature* yang berpengaruh terhadap proses kategorisasi. Kemudian diamati nilai *precision* dan *recall* menggunakan algoritma naïve bayess yang terdapat pada tools WEKA. Setelah dilakukan percobaan dengan menetapkan nilai *threshold* untuk pemilihan *feature*, menunjukkan semakin tinggi nilai *threshold* yang ditetapkan maka jumlah *feature* yang dipilih oleh sistem semakin sedikit, namun performansi hasil kategorisasi meningkat. Performansi kategorisasi mencapai nilai tertinggi ketika ditetapkan *threshold* 0.95. Kemudian dilakukan perbandingan akurasi antara dataset sebelum dilakukan *feature selection* dan dataset setelah dilakukan *feature selection*, menunjukkan akurasi yang dihasilkan setelah dilakukan *feature selection* lebih tinggi daripada dataset sebelum dilakukan *feature selection*.

Kata kunci : *feature selection*, kategorisasi teks, *Ambiguity Measure*