

ABSTRAKSI

Adanya pertumbuhan yang pesat dari informasi pada masa kini, menjadikan kategorisasi teks (*text categorization*) yang merupakan suatu proses pengklasifikasian dokumen ke dalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau ke dalam kelas-kelas dari dokumen-dokumen yang sama, sebagai suatu teknik kunci dalam penanganan dan pengorganisasian data yang berupa teks.

Salah satu metode klasifikasi yang digunakan adalah *Support Vector Machine* (SVM). Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada *input space*. SVM secara konsep merupakan *linear classifier*, tetapi SVM dapat dimodifikasi dengan menggunakan *kernel trick*, sehingga bisa digunakan untuk menyelesaikan permasalahan non-linear. Walaupun masih terbilang metode baru, SVM memberikan performansi yang lebih baik dibandingkan metode-metode *learning* lainnya, seperti Naïve Bayesian, Bayesian Network, Decision Trees[4], atau Artificial Neural Network[1].

Melalui serangkaian pengujian terhadap sejumlah dokumen yang telah diolah menjadi suatu representasi data berupa matriks vektor, SVM berhasil mengklasifikasikan dokumen dalam jumlah besar (>3000 dokumen) ke dalam kategorinya dengan tingkat akurasi yang memuaskan, yaitu di atas 80%.

Kata Kunci : kategorisasi teks, dokumen, *Support Vector Machine*