

ABSTRAK

Klasifikasi dokumen teks adalah masalah yang sederhana namun sangat penting karena manfaatnya cukup besar mengingat jumlah dokumen yang ada setiap hari semakin bertambah. Namun, kebanyakan teknik klasifikasi dokumen yang ada memerlukan *labeled documents* dalam jumlah besar untuk melakukan tahap *training* dan *testing*. Dalam melakukan klasifikasi dokumen, pada tugas akhir ini digunakan algoritma *Principal Component Analysis* yang dikombinasikan dengan *Support Vector Machines* untuk supervised document. *Principal Component Analysis* merupakan suatu teknik yang dapat digunakan untuk mengekstraksi struktur dari suatu data yang berdimensi tinggi tanpa menghilangkan informasi yang signifikan pada keseluruhan data. Kemudian dibutuhkan sebuah algoritma yang dapat menghasilkan prediksi dan akurasi dari dokumen tersebut yaitu *Support Vector Machines* (SVM). SVM adalah metode *learning machine* yang bekerja atas prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya.

Hasil dari pengujian sistem menggunakan data yang direduksi oleh *Principal Component Analysis* (PCA) memiliki akurasi yang sedikit lebih rendah untuk *dataset* tertentu dibandingkan tanpa menggunakan PCA. Data yang digunakan adalah data *R8 of Reuters-21578 Text Categorization Collection Data Set*. Akurasi terbaik pada penelitian ini dihasilkan dari metode SVM dengan akurasi rata-rata 98.95%, sedangkan untuk metode SVM + PCA akurasi yang diperoleh rata-rata 96.7866%.

Kata kunci :

Klasifikasi Dokumen, *Principal Component Analysis*, *Support Vector Machine*