

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 *Naïve Bayes* 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 *Naïve Bayes*. *Naive bayesian* klasifikasi adalah suatu klasifikasi berpeluang sederhana berdasarkan aplikasi teorema Bayes dengan asumsi antar variabel penjelas saling bebas (*independen*). Dalam hal ini, diasumsikan bahwa kehadiran atau ketiadaan dari suatu kejadian tertentu dari suatu kelompok tidak berhubungan dengan kehadiran atau ketiadaan dari kejadian lainnya. Hasil dari pengujian sistem menghasilkan data yang direduksi oleh *Principal Component Analysis* (PCA) memiliki akurasi yang sama untuk *dataset* tertentu dibandingkan tanpa menggunakan PCA.

Data yang digunakan merupakan R8 data from reuter is R8 data from Reuters-21578 Text Categorization of Collection Data Set. Hasil akurasi tertinggi yang diperoleh dari klasifikasi menggunakan *Naïve Bayes* terdapat pada kombinasi no 1 yaitu 59,111% dan untuk *Naïve Bayes* + PCA didapatkan pada kombinasi 10 dengan akurasi 64%.

Kata kunci :

Klasifikasi Dokumen, *Principal Component Analysis*, *Naïve Bayes*