Abstrak

Perkembangan teknologi menyebabkan terjadinya penumpukan data berupa dokumen teks baik secara *online* maupun *offline*. Dokumen teks yang menumpuk menyebabkan sulitnya mencari dokumen yang sesuai dengan kebutuhan. Untuk kemudahan pencarian dokumen yang sesuai, maka dibutuhkan kata kunci yang mengiringi dokumen. Kata kunci mewakili isi dokumen secara keseluruhan, sehingga pembaca dapat dengan mudah mencari dokumen sesuai dengan yang dibutuhkan.

Pada tugas akhir ini, diimplementasikan metode Naive Bayes untuk mengekstraksi kata kunci dokumen. Proses ini membutuhkan dokumen *training* yang telah disertai kata kunci, sehingga dengan menggunakan fitur-fitur tertentu dapat memberikan *learning* kepada sistem bagaimana probabilitas atau peluang kata pada dokumen masukan menjadi kata kunci.

Pengujian dilakukan untuk mengetahui keakurasian kata kunci yang dihasilkan oleh sistem dengan menggunakan metode Naive Bayes berdasarkan parameter *precision, recall,* dan *f-measure.* Penambahan jumlah dokumen *training*, menyebabkan meningkatnya keakurasian hasil ekstraksi kata kunci. Namun, penggunaan jumlah dokumen *training* yang terlalu besar menyebabkan penurunan nilai keakurasian. Penggunaan 4 fitur pada proses ekstraksi memberikan hasil keakurasian yang lebih baik dibandingkan dengan penggunaan 2 fitur. Kemampuan Naive Bayes dalam mengekstraksi kata kunci dengan benar dapat dilihat dari nilai *recall*. Dalam rentang 20 jumlah kata kunci yang dihasilkan sistem, Naive Bayes mampu memberikan nilai *recall* sebesar 0,75 (sekitar 75% sistem mampu mengekstraksi benar kata kunci yang sesuai dengan dokumen masukan). Ekstraksi kata kunci dengan menambahkan eliminasi *stopwords* memberikan hasil keakurasian yang lebih baik dibanding dengan tanpa eliminasi *stopwords*.

Kata Kunci : Ekstraksi kata kunci, kata kunci, naive bayes