

## ABSTRAK

Pada sekarang ini ketersediaan informasi berbentuk dokumen teks sebagian besar sudah berbentuk elektronik (*softcopy*). Penyimpanan media teks tersebut akan mengalami perkembangan yang sangat besar pada masa yang mendatang. Yang perlu dilakukan adalah penggolongan dokumen yang berada dalam satu kumpulan dokumen (*corpus*) ke dalam kategori yang sesuai dengan isi dokumen tersebut. Pengklasifikasian dokumen sulit dilakukan jika menggunakan *Query* biasa, maka hasil kurang spesifik dapat mengakibatkan membanjirnya beberapa dokumen yang tidak relevan. *Text Mining* adalah suatu bidang satu bidang khusus dari *Data Mining* yang memberikan solusi dari permasalahan seperti pemrosesan, pengelompokkan dan menganalisis *unstructured text* dalam jumlah besar. *Feature selection* adalah suatu bentuk upaya peningkatan algoritma pembelajaran yang digunakan untuk menggolongkan dokumen ke dalam kategori tertentu dengan cara menemukan suatu bentuk pola yang relevan. *Chi Squared* adalah salah satu metode yang digunakan untuk proses *Feature Selection*. Sedangkan metode klasifikasi dokumen yang digunakan adalah metode *Naïve Bayes Classifier* (NBC) yang digunakan untuk memecahkan masalah berhubungan dengan proses klasifikasi.

*Kata Kunci : Corpus, Klasifikasi, Data Mining, Text Mining, Query, Feature Selection, Chi Squared, Naïve Bayes Classifier.*