

Abstrak

Salah satu *task* Data Mining yaitu klasifikasi teks yang merupakan suatu cara untuk mengelompokkan dokumen kedalam topik atau kategori tertentu. Saat ini klasifikasi teks dapat digunakan untuk mengklasifikasikan sebuah teks yang mengandung kata berhomograf. Dengan adanya klasifikasi teks ini mempermudah pengguna dalam menentukan lafal sebuah kata yang homograf.

Metode Naive Bayes dan *K-Nearest Neighbor* merupakan dua dari beberapa metode klasifikasi. Dua metode ini belum bisa dipastikan akurasi sebelum dilakukan perbandingan secara langsung dan dengan data yang sama. Metode Naive Bayes merepresentasikan dokumen kedalam bentuk *probability model*. *Probability model* tersebut yang digunakan sebagai model untuk mengelompokkan suatu dokumen. Sedangkan metode *K-Nearest Neighbor* menyimpan sekelompok data *training set* sebagai acuan untuk mengelompokkan suatu dokumen.

Pada Tugas Akhir ini menggunakan metode Naive Bayes dan *k-nearest neighbor* dalam kasus klasifikasi teks yang mengandung kata homograf. Selain itu dilakukan juga analisis terhadap performansi hasil dari klasifikasi dua metode ini. Parameter performansi yang digunakan adalah *precision*, *recall*, dan *f-measure*. Pada akhir pengujian metode naïve bayes menghasilkan nilai rata-rata *f-measure* yang lebih tinggi dibanding *k-nearest neighbor*.

Kata kunci: *probability model, training set, precision, recall, f-measure*