

Abstrak

Salah satu permasalahan dalam bidang kategori dokumen dan bioinformatik adalah karakteristik data yang mempunyai lebih dari satu label (*multi-label*). Permasalahan tersebut dapat diselesaikan dengan mengelompokkan data menggunakan metode klasifikasi. Dalam klasifikasi *multi-label*, setiap data dalam *training set* dipasangkan dengan label-label[4]. *Training set* tersebut akan di-*training* dengan menggunakan salah satu metode klasifikasi (*classifier*) tertentu, sehingga dihasilkan pemodelan yang dapat digunakan untuk memprediksi label-label dari data yang belum diketahui label-labelnya (*test set*).

Untuk menyelesaikan permasalahan tersebut maka *classifier* yang digunakan untuk klasifikasi data *multi-label* adalah *Support Vector Machine (SVM)* dan *k-Nearest Neighbor (k-NN)*. Kedua metode tersebut berbasis statistik dengan tingkat akurasi yang cukup tinggi dalam beberapa penelitian[4,10,7]. *SVM* merupakan metode *machine learning* yang selalu berusaha menemukan *hyperplane* terbaik untuk memisahkan kelas pada *input space*. Sedangkan *k-NN* merupakan metode *instance-base* yang dikenal sebagai “*lazy learning*”, karena mengklasifikasi data berdasarkan "k" tetangga terdekat data dan menghitung *maximum a posterior (MAP)* untuk menentukan label-label dari *test set*[7]. Efektivitas dari kedua metode tersebut diukur dengan menggunakan matriks evaluasi yang meliputi *accuracy*, *precision*, *recall*, *hammingLoss*, *one-error*, *rankingloss*, dan *coverage*.

Pada tugas akhir ini, analisis dilakukan dengan membandingkan hasil perhitungan matriks evaluasi dari kedua *classifier* tersebut untuk mengetahui *classifier* yang paling handal dalam mengklasifikasikan data *multi-label*. Selain itu, analisis juga dilakukan dengan membandingkan antara *classifier* yang khusus untuk *multi-label* dengan *classifier* yang umum.

Kata kunci : klasifikasi *multi-label*, *classifier*, *SVM*, *k-NN*, *hyperplane*, *maximum a posterior*, matriks evaluasi