

Klasifikasi *Multi-label* pada Topik Berita Berbahasa Indonesia Menggunakan *K-Nearest Neighbor*

Nikmah Isnaini¹, Adiwijaya², Mohamad Syahrul Mubarak³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹imaisna@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³sym.milo@gmail.com

Abstrak

Berita telah menjadi kebutuhan pokok manusia seiring dengan berkembangnya teknologi dan internet. Perkembangan teknologi dan internet ini menyebabkan proses pendistribusian informasi pada berita beralih dari cara penyampaian era media cetak menuju era digital. Masalah lain yang muncul ketika mengklasifikasi berita adalah *multi-label*. Klasifikasi *multi-label* berbeda dengan klasifikasi *single label*. Klasifikasi *single label* akan mengklasifikasikan suatu dokumen ke dalam satu label saja. Sedangkan klasifikasi *multi-label* dapat mengelompokkan suatu dokumen ke lebih dari satu label. Sebagai contoh, sebuah artikel berita yang di dalamnya membahas secara rinci mengenai deteksi dini kanker ovarium dengan pendekatan bioinformatika, dapat memiliki lebih dari satu label yaitu kesehatan, bioinformatika, dan wanita. Pada tugas akhir ini, dibangun sebuah model klasifikasi yang mampu mengidentifikasi kelas pada masing-masing artikel berita secara *multi-label* dengan menggunakan metode *K-Nearest Neighbor*. Kelebihan *K-Nearest Neighbor* adalah algoritma yang sangat cocok untuk kasus *multi-label*, bahkan KNN bisa lebih unggul dibanding *classifier* yang lain. Dari sistem yang dibuat, didapatkan hasil nilai performa sistem yang diukur dengan *proximity measure* yaitu perbandingan antara *Manhattan Distance*, *Euclidean Distance*, dan *Supremum Distance* dengan menggunakan parameter $K=11$, menghasilkan nilai *Hamming Loss* sebesar 11,16%.

Kata kunci : klasifikasi, *multi-label*, *K-Nearest Neighbor*, *proximity measure*, *Hamming Loss*.
