

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

Abstract

News has become a basic human need along with the development of technology and the internet. The development of technology and the internet has led to the process of distributing information on the news shifting from the way of delivering the print media era to the digital era. Another problem that arises when classifying news is multi-label. The multi-label classification differs from the single label classification. Single label classification will classify a document into one label only. While multi-label classification can group a document to more than one label. For example, a news article in which to discuss in detail the early detection of ovarian cancer with bioinformatics approach, can have more than one label of health, bioinformatics, and women. In this final project, a classification model was developed that was able to identify the classes in each news article multi-labeled using K-Nearest Neighbor method. Excess K-Nearest Neighbor is an algorithm that is very suitable for multi-label cases, even KNN can be superior to other classifier. From the system that is made, the result of system performance value which measured with proximity measure that is comparison between Manhattan Distance, Euclidean Distance and Supremum Distance using parameter $K = 11$, yields Hamming Loss value 11,16%.

Keywords: classification, multi-label, K-Nearest Neighbor, proximity measure, Hamming Loss.
