Kategorisasi Berita *Multi-Label* Berbahasa Indonesia Menggunakan Algoritma *Random Forest*

Brama Hendra Mahendra[1], Adiwijaya[2], Untari Novia Wisesty[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]bramahendramahendra@students.telkomuniversity.ac.id, [2]adiwijaya@telkomuniversity.ac.id,
[3]untarinw@staff.telkomuniversity.ac.id

Abstract
News is an information about something that is happening or has happened. Along with the development of technology where news is presented in the form of websites, it causes a large number of digital news released by several news portals issued. From the abundance of news document that exist, it has an impact on the number of news document that have the same meaning. Based on the description above, it needs a good news categorization methods to facilitate information retrieval. In this case, there are many methods can be done in categorizing the news one of them by the Random Forest method. But before using this method, there are several steps must be taken to determine the value of random forest. One step that must be done is to determine feature extraction using the Regular Expression method and complete it by weighting of TF-IDF, and then doing Cross-Validation. The data used consists of two types, testing data and training data. For the results of the trial obtained a value of 0.126 from the Hamming Loss equation process.

Keywords: Random Forest, News Document, Feature Extraction, Weighting of TF-IDF, Cross-Validation, Hamming Loss