

Klasifikasi Multi Label Pada Hadis Bukhari Terjemahan Bahasa Indonesia menggunakan Random Forest, Mutual Information, dan Chi-Square

Hadiyan Nadzri Harish¹, Said Al Faraby², Mahendra Dwifabri³

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

¹Hnadzriaji@student.telkomuniversity.ac.id, ²saidalfaraby@telkomuniversity.ac.id,

³mahendradp@telkomuniversity.ac.id

Abstract

Hadith is the second source of law for Muslims after the Qur'an. Therefore, Muslims are encouraged to know and practice it. Generally, Hadith is collected by several high priests, one of which is the Hadith collection of Imam Bukhari. There are several categories of teachings in the Hadith, such as the types of containing suggestions, prohibitions, and information. To identify the characteristics of Hadith translated into Indonesian by category, this research will build a system capable of classifying Hadith into three categories, namely recommendations, prohibitions, and information. In classifying Hadith, a classification system is needed. In developing this system, various classification methods can use, one of which is Random Forest. Random forest is a classification method that can generalize high-dimensional data. This Random Forest is believed to complete the classification process with accurate results but has a weakness, namely overfitting when dealing with data types with many features. In this study, the feature selection that will use is Chi-Square. The feature selection method can help select features from the original feature set to eliminate are features not relevant to each class. Optimum results obtained from several tests in this study showed an accuracy value of 91.7% of data classified correctly using Chi-Square as a selection feature, and without stemming process.

Keywords : Multi-label classification, bukhari's hadith, Random forest, Mutual Information, Chi-Square.
