
Multilabel Classification of Qur'anic Verse Topics Using Random Forest and Naïve Bayes

Imran Zulkarnaen¹, Kemas Muslim Lhaksana²

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

¹imranzulkarnaen@students.telkomuniversity.ac.id,

²kemasmuslim@telkomuniversity.ac.id

Abstract

The Qur'an, as the holy book of Muslims, holds profound meanings, covering aspects of faith, worship, and social ethics. However, the complexity of the language in the Qur'an poses challenges in grouping its verses into specific thematic categories, especially with traditional approaches that often cannot explore the semantic relationships between words in depth. To address this challenge, this research develops a graph mining-based multilabel classification system, utilizing centrality measures. The system involves creating a word graph to represent the relationship between words, and applying random forest and naïve bayes algorithms to classify Qur'anic verses into eight thematic categories. Data processing includes stopwords removal, tokenization, and feature extraction based on centrality, such as closeness, betweenness, and eigenvector. The results show that the use of betweenness centrality with the use of stopwords provides the best performance, with a Hamming loss value of 0.1631 in random forest. These findings emphasize the superiority of the graph-based approach in understanding complex relationships between words in the Qur'anic text and contribute to the development of more efficient technology-based thematic classification methods.

Keyword: Multilabel classification, Thematic, Al-Qur'an, Graph, Centrality, Graph Mining, Hamming Loss.
