
Klasifikasi *Multilabel* pada Topik Ayat Al-Qur'an Menggunakan *Random Forest* dan *Naïve Bayes*

Imran Zulkarnaen¹, Kemas Muslim Lhaksana²

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

¹imranzulkarnaen@students.telkomuniversity.ac.id,

²kemasmuslim@telkomuniversity.ac.id

Abstrak

Al-Qur'an, sebagai kitab suci umat Islam, menyimpan makna yang mendalam, mencakup aspek akidah, ibadah, dan etika sosial. Namun, kerumitan bahasa dalam Al-Qur'an menimbulkan tantangan dalam pengelompokan ayat-ayatnya ke dalam kategori tematik tertentu, terutama dengan pendekatan tradisional yang sering kali tidak dapat menggali hubungan semantik antar kata secara mendalam. Untuk mengatasi tantangan ini, penelitian ini mengembangkan sistem klasifikasi multilabel yang berbasis *graph mining*, dengan memanfaatkan pengukuran *centrality*. Sistem tersebut melibatkan pembuatan graf kata untuk merepresentasikan hubungan antar kata, serta penerapan algoritma *random forest* dan *naïve bayes* dalam mengklasifikasikan ayat-ayat Al-Qur'an ke dalam delapan kategori tematik. Proses pengolahan data mencakup penghapusan kata henti (*stopwords*), tokenisasi, dan ekstraksi fitur berdasarkan *centrality*, seperti *closeness*, *betweenness*, dan *eigenvector*. Hasil penelitian menunjukkan bahwa penggunaan *betweenness centrality* dengan penggunaan kata henti memberikan performa terbaik, dengan nilai *Hamming loss* sebesar 0.1631 pada *random forest*. Temuan ini menekankan keunggulan pendekatan berbasis graf dalam memahami hubungan kompleks antar kata dalam teks Al-Qur'an serta berkontribusi pada pengembangan metode klasifikasi tematik berbasis teknologi yang lebih efisien.

Kata kunci: klasifikasi *Multilabel*, Tematik, Al-Qur'an, Graf, Sentralitas, *Graph Mining*, *Hamming Loss*.
