

Abstrak

Dilihat dari banyaknya informasi yang terkandung dalam Al-Qur'an, akan sulit untuk menggali informasi tersebut dengan cara manual, terlebih untuk orang yang ingin memahami lebih dalam terkait isi Al-Qur'an. Oleh karena itu, dibutuhkan satu cara untuk menemukan informasi terkait topik tertentu yang telah diklasifikasi dalam Al-Qur'an, terlebih dalam satu ayat Al-Qur'an bisa jadi memiliki lebih dari satu topik bahasan (multilabel). Dalam penelitian ini diteliti bagaimana membangun classifier untuk menyelesaikan masalah klasifikasi multilabel pada topik ayat Al-Qur'an dengan metode k-Nearest Neighbor. Hasil pengujian dengan mencari nilai k paling optimal, mendapatkan hasil k=25 dengan nilai rata-rata *hamming loss*=0.134875. Dalam penelitian ini juga dilakukan perbandingan antara ekstraksi fitur *Weighted TF-IDF* dan *TF-IDF*. Hasil yang didapatkan dari perbandingan tersebut bahwa *Weighted TF-IDF* lebih baik dibanding *TF-IDF*. Untuk pengujian pengaruh *stopword removal* dan *lemmatization* dengan nilai k optimal, pada kasus tanpa *stopword removal* hasil yang didapat adalah 0.136375, sedangkan tanpa *lemmatization* performa yang didapat adalah 0.13537. Untuk kasus pengujian tanpa kedua proses tersebut didapatkan hasil nilai rata-rata *hamming loss* sebesar 0.1373125.

Kata kunci: Al-Qur'an, klasifikasi, *multilabel*, KNN, *hamming loss*