

Abstrak

Hadis merupakan sumber hukum kedua bagi umat Islam setelah Al-Quran. Banyak sekali hadis yang telah diriwayatkan, namun Hadis Bukhari memiliki tingkat kesahihan paling tinggi menurut para ulama. Seiring dengan perkembangan teknologi, hadis sangatlah mudah didapatkan melalui dunia digital. Akan tetapi untuk mempelajari hadis tidak semudah yang kita bayangkan. Banyaknya hadis yang ada dan juga belum dikategorikan membuat belajar hadis dengan kategori tertentu sangat sulit dilakukan. Oleh sebab itu penulis melakukan penelitian klasifikasi anjuran, larangan dan informasi pada Hadis Sahih Al-Bukhari terjemahan Bahasa Indonesia yang diharapkan dapat mempermudah masyarakat dalam mempelajari hadis. Proses klasifikasi menggunakan model unigram/bigram dengan *Mutual Information* (MI) sebagai seleksi fitur dan *Support Vector Machine* (SVM) sebagai metode klasifikasi. Pada penelitian ini dilakukan beberapa skenario pengujian dengan memodifikasi *term model*, *preprocessing*, *feature selection* dan menggunakan beberapa metode klasifikasi untuk membuktikan bahwa SVM merupakan salah satu metode klasifikasi teks yang cocok digunakan. Pengujian dengan menggunakan model unigram, tidak menggunakan stopword/stemming, menggunakan MI dan menggunakan SVM memberikan nilai hamming loss terbaik yaitu 0.0686. Hasil penelitian yang diperoleh juga menunjukkan bahwa metode SVM dengan menggunakan MI lebih baik daripada metode klasifikasi teks yang lain.

Kata kunci: Hadis Bukhari, *Hamming loss*, Klasifikasi, *Mutual Information*, *Preprocessing*, *Support Vector Machine*.