

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

Hadiyan Nadzri Harish<sup>1</sup>, Said Al Faraby<sup>2</sup>, Mahendra Dwifabri<sup>3</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>Hnadzriaji@student.telkomuniversity.ac.id, <sup>2</sup>saidalfaraby@telkomuniversity.ac.id,

<sup>3</sup>mahendradp@telkomuniversity.ac.id

---

## Abstrak

Hadis merupakan sumber hukum kedua bagi umat Islam setelah Al-Qur'an. Oleh karena itu, umat Islam dianjurkan untuk mengetahui dan mengamalkannya. Umumnya Hadis dikoleksi oleh beberapa imam besar, salah satunya adalah koleksi Hadis Imam Bukhari. Dalam Hadis terdapat beberapa kategori jenis ajaran, seperti jenis ajaran Hadis yang mengandung anjuran, larangan, dan informasi. Untuk mengenali karakteristik Hadis terjemahan Bahasa Indonesia berdasarkan kategorinya, pada penelitian ini akan dibangun sebuah sistem yang mampu mengelompokkan Hadis kedalam tiga kategori yaitu anjuran, larangan, dan informasi. Dalam mengelompokkannya diperlukan sebuah sistem klasifikasi. Pada penelitian ini, berbagai metode klasifikasi dapat digunakan salah satunya adalah Random Forest. Random forest merupakan metode klasifikasi yang memiliki kemampuan menggeneralisasi suatu data berdimensi tinggi. Random Forest ini dipercaya dapat menyelesaikan proses klasifikasi dengan hasil yang akurat, namun memiliki kelemahan yaitu terjadinya *overfitting* ketika menghadapi jenis data dengan jumlah fitur yang banyak. Dalam penelitian ini seleksi fitur yang akan digunakan yaitu Chi-Square. Metode seleksi fitur dapat membantu proses penyeleksian subset fitur dari sekumpulan fitur asli dengan tujuan mengurangi jumlah fitur yang tidak relevan terhadap masing-masing kelas. Hasil Optimum yang didapat dari beberapa pengujian pada penelitian ini menunjukkan nilai akurasi sebesar 91,7% data terklasifikasi dengan benar dan menggunakan Chi-Square sebagai fitur seleksi, namun tanpa stemming.

**Kata kunci:** klasifikasi multi-label, hadis bukhari, Random Forest, Mutual Information, Chi-square

---