

# Kategorisasi Berita *Multi-Label* Berbahasa Indonesia Menggunakan Algoritma *Random Forest*

Brama Hendra Mahendra<sup>1</sup>, Adiwijaya<sup>2</sup>, Untari Novia Wisesty<sup>3</sup>

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

<sup>1</sup>bramahendramahendra@students.telkomuniversity.ac.id, <sup>2</sup>adiwijaya@telkomuniversity.ac.id,

<sup>3</sup>untarinw@staff.telkomuniversity.ac.id

---

## Abstrak

Berita merupakan informasi mengenai sesuatu yang sedang terjadi atau sudah terjadi. Seiring dengan berkembangnya teknologi dimana berita disajikan dalam bentuk *website* karena hal itu menyebabkan jumlah berita digital yang dirilis oleh beberapa portal berita setiap harinya menjadi sangat banyak. Dari banyaknya ketersediaan dokumen berita yang ada, berdampak pada banyaknya dokumen berita yang memiliki makna yang sama. Berdasarkan dari uraian diatas dibutuhkan metode-metode pengkategorian berita yang baik untuk memudahkan dalam pengambilan informasi. Dalam hal ini, banyak metode yang dapat dilakukan dalam mengkategorikan berita salah satunya dengan metode *Random Forest*. Tapi sebelum menggunakan metode tersebut, terdapat beberapa langkah yang harus dilakukan dalam menentukan nilai dari *random forest*. Salah satu langkah yang harus dilakukan adalah menentukan *feature extraction* dengan metode *Regular Expression* dan dilanjutkan dengan pembobotan *TF-IDF* lalu setelah itu melakukan *Cross-Validation* dengan *k-Fold*. Dataset yang digunakan terdiri dari dua jenis yaitu data *testing* dan data *training*. Untuk hasil uji coba diperoleh nilai sebesar 0,126 dari proses persamaan *Hamming Loss*.

**Kata kunci:** Random Forest, Dokumen Berita, Feature Extraction, Pembobotan TF-IDF, Cross-Validation, Hamming Loss

---