

Klasifikasi Berita Bahasa Indonesia Menggunakan *Mutual Information* dan *Support Vector Machine*

Lalu Gias Irham¹, Adiwijaya², Untari Novia Wisesty³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹giasirham@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³untarinw@telkomuniversity.ac.id

Abstrak

Berita merupakan sumber informasi yang disebarkan dalam berbagai macam media. Agar memudahkan pembaca berita dalam memperoleh berita yang diinginkan, maka berita perlu diklasifikasikan. Banyaknya jumlah berita yang tersebar menimbulkan kesulitan dalam mengelompokkan berita tersebut berdasarkan topiknya. Oleh sebab itu penulis melakukan penelitian untuk mengklasifikasikan berita ke dalam 12 kelas (budaya, ekonomi, hiburan, hukum, kesehatan, gaya hidup, otomotif, pendidikan, politik, olahraga, teknologi, dan wisata) secara otomatis terhadap 360 data berita Bahasa Indonesia. Pada penelitian ini dilakukan beberapa scenario pengujian untuk melihat pengaruh dari metode *stopword removal* dan *stemming* pada *preprocessing* data, pengaruh *mutual information* dalam menyeleksi fitur, dan performansi *Support Vector Machine* dalam mengklasifikasikan data berita. Hasil pengujian menunjukkan data yang hanya menggunakan *stemming* tanpa *stopword removal*, menggunakan fitur seleksi MI dan metode klasifikasi SVM menghasilkan hasil terbaik yaitu 94.24%, dibandingkan dengan metode yang lainnya.

Kata kunci : berita, klasifikasi teks, *support vector machine*, seleksi fitur, *mutual information*