

INTISARI

KLASIFIKASI BERITA ONLINE BERDASARKAN JUDUL DAN LEAD MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE*

Oleh

Valery Cindy Claudia

14/364136/PA/15912

Seiring berjalannya waktu, informasi yang berupa dokumen, teks atau artikel semakin banyak dihasilkan dan dibutuhkan oleh seluruh kalangan masyarakat. Karena perkembangan teknologi, informasi disampaikan menggunakan sistem berbasis web secara update yang biasa disebut berita *online*. Berita *online* yang jumlahnya semakin banyak perlu diklasifikasikan agar mempermudah pembaca untuk menemukan berita pada kategori yang lebih spesifik. Penelitian ini mengklasifikasikan berita olahraga berdasarkan 11 sub kategori yaitu atletik, basket, bulutangkis, formula 1, moto GP, renang, sepakbola, tenis, tinju, olahraga lain, dan non olahraga.

Klasifikasi berita *online* berkategori olahraga pada penelitian ini dilakukan dengan algoritma *support vector machine*. Judul dan *lead* berita menjadi bagian spesifik dari berita yang diproses karena sudah mengandung inti dari berita itu sendiri tanpa harus memproses bagian yang lain. Penelitian diawali dengan proses *scraping* untuk pengumpulan data, labeling, *preprocessing*, seleksi fitur dengan pembobotan TF-IDF, kemudian klasifikasi dan evaluasi menggunakan *K-fold cross validation*.

Klasifikasi berita *online* berkategori olahraga dengan memproses judul dan *lead* dari 1117 berita menghasilkan nilai rata-rata akurasi, presisi, recall dan f-measure sebesar 94,3%, 95,3%, 94,3% dan 94,4% menggunakan *10-fold cross validation* dan algoritma *support vector machine* saat jumlah fitur maksimum sebesar 2000.

Kata Kunci: *klasifikasi berita, support vector machine, berita online*

ABSTRACT

ONLINE NEWS CLASSIFICATION BASED ON TITLE AND LEAD USING SUPPORT VECTOR MACHINE ALGORITHM

by

Valery Cindy Claudia

14/364136/PA/15912

Over time, information in the form of documents, text or articles more and more generated and required by the whole society. Due to technological developments, information is delivered using an updated web-based system commonly called online news. An increasing number of online news needs to be classified to make it easier for readers to find news in more specific categories. This study classifies sports news based on 11 sub categories: athletics, basketball, badminton, formula 1, moto GP, swimming, football, tennis, boxing, other sports, and non sports.

The classification of sports news online in this research is done with support vector machine algorithm. Title and news leads become a specific part of the news processed because it already contains the core of the news itself without having to process the other part. The research begins with a scraping process for data collection, labeling, preprocessing, feature selection with TF-IDF weighting, then classification and evaluation using K-fold cross validation.

The classification of sports online news by processing titles and leads from 1117 news resulted in average values of accuracy, precision, recall and f-measure of 94.3%, 95.3%, 94.3% and 94.4% using 10-fold cross validation and the support vector machine algorithm when the maximum number of features is 2000.

Keywords: *News classification, support vector machine, online news*