
Pengembangan Model Pengklasifikasi Naïve Bayes untuk Seleksi Penelusuran URL Halaman Detail Informasi Produk Tas Wanita pada Situs X

Citra Lestari Teknik Informatika Universitas Ciputra, UC Town CitraRaya, Surabaya 60219

ABSTRAK

Pencarian informasi dengan mesin pencari telah umum dilakukan. Umumnya, untuk informasi tentang detail sebuah produk, pengguna mencari pada situs-situs yang menjual produk tersebut. Selain mesin pencari seperti Google yang bersifat universal, sebenarnya terdapat mesin pencari yang menelusuri web tertentu saja untuk kebutuhan yang sangat spesifik. Mesin seperti ini melakukan penelusuran secara sekaligus sehingga membutuhkan biaya, waktu dan ruang, yang besar, sehingga perlu dilakukan seleksi halaman yang hendak ditelusuri dan unduh. Karya ilmiah ini membuat sebuah model pengklasifikasi menggunakan Naïve Bayes. Model pengklasifikasi ini digunakan untuk seleksi URL halaman yang akan ditelusuri dan diunduh oleh penelusur web. Halaman yang diinginkan adalah halaman detail informasi tas wanita pada situs X, salah satu situs toko online terbesar di Indonesia. Dengan pendekatan klasifikasi teks, maka sebuah URL halaman dianggap sebagai dokumen. Dokumen atau URL direpresentasikan dalam model Boolean, yang melihat muncul atau tidaknya sebuah istilah pada suatu URL. Kumpulan dokumen dilabeli sebagai tas wanita (*bw*) atau bukan tas wanita (*nbw*). Model dibangun dengan melatih 800 dokumen. Model ini menemukan istilah “bags” sebagai istilah dengan probabilitas posterior tertinggi (0.99), sedangkan istilah “bag” meskipun lebih kerap muncul namun memiliki kekuatan yang sama pada kedua kelas (probabilitas posterior pada kelas *bw* = 0.57) sehingga dapat menyebabkan bias pada hasil klasifikasi. Model pengklasifikasi ini kemudian diuji menggunakan 325 dokumen yang berbeda dengan data dokumen latih. Akurasi dari pengujian tersebut adalah sebesar 93,2%.

Kata kunci: model pengklasifikasi, Naïve Bayes, penelusuran web, tas wanita, seleksi url

1. Pendahuluan

Pencarian informasi kini telah umum dilakukan dengan memanfaatkan teknologi internet. Informasi tersebut dapat diperoleh dari situs, baik komersial maupun non-komersial. Umumnya, untuk informasi tentang detail sebuah produk, seperti merek, spesifikasi, ukuran, harga, pengguna internet mencarinya pada situs-situs yang menjual produk tersebut. Pencarian informasi secara manual biasanya dilakukan dengan bantuan mesin pencari, seperti Google. Pengguna menuliskan kata kunci dari informasi yang ingin diketahui kemudian mesin pencari akan memberikan daftar situs yang sesuai.

Mesin pencari seperti Google bersifat universal. Mesin pencari seperti ini menelusuri dan menyimpan semua situs secara berkala (Menczer, 2011). Untuk tujuan yang lebih spesifik dan khusus, dapat dibuat sebuah mesin pencari bertipe topikal yang hanya menelusuri beberapa website tertentu dan hanya menyimpan halaman tertentu, misalnya halaman detail informasi produk.

Tidak seperti penelusur Universal yang mengunjungi secara berangsur, mesin pencari bertipe topikal menelusuri halaman-halaman sekaligus pada satu waktu. Jika semua halaman ditelusuri dan diunduh, tentunya mesin pencari perlu memiliki memori yang cukup besar (Menczer, 2011). Alternatif lain adalah membuat mesin pencari yang