

Abstrak

Selama beberapa tahun terakhir, World Wide Web telah menjadi salah satu sumber penting untuk pencarian informasi. Oleh karena itulah, seiring dengan jumlah web yang terus meningkat, kemampuan untuk mendapatkan informasi yang spesifik dan relevan adalah sama pentingnya dengan kehandalan web itu sendiri untuk menampilkan informasi sesuai dengan masing-masing kelasnya. Pada dasarnya, halaman web merupakan sebuah *hypertext*. Selain teks dan komponen multimedia lainnya, web juga terdiri dari konten lainnya yang berupa *hyperlink*, *tag* HTML dan meta data. Klasifikasi halaman web menjadi tidak mudah mengingat atribut dalam dokumen HTML berupa teks yang jumlahnya lebih banyak ditemui dibandingkan dengan komponen media lainnya seperti *audio*, *video*, atau *image*.

Permasalahan klasifikasi tersebut diselesaikan dengan menerapkan algoritma *multinomial naïve bayes classifier* dan dengan memanfaatkan komponen halaman web yang mengandung *meta tag description* dan *meta tag keywords*. Kemudian dari penerapannya, dianalisis ketepatan kelas hasil klasifikasi dan tingkat akurasi prediksi serta dianalisis faktor-faktor yang mempengaruhi tingkat akurasi hasil klasifikasi tersebut.

Multinomial naïve bayes classifier bekerja berdasarkan sekumpulan bukti (*evidence*) dan kelas (*class*). Dengan melakukan pelatihan (*training*) terhadap sebagian data sampel, dapat ditentukan probabilitas kepastian (*likelihood probability*) dari sebuah bukti jika diberikan kelas tertentu. *Multinomial naïve bayes classifier* juga menggunakan probabilitas sebelumnya (*prior probability*) dari sebuah kelas, yang perhitungannya dapat didasarkan dari sampel data tersebut. Dari analisis sampel data tersebut, jika diberikan sebuah dokumen baru yang terdiri dari sekumpulan bukti, probabilitas setiap kelas terhadap dokumen tersebut (*posterior probability*) dapat ditentukan.

Kata Kunci: halaman web, meta tag, klasifikasi, multinomial naïve bayes classifier