

Pengklasifikasian artikel berita berbahasa Indonesia secara otomatis menggunakan naive bayes classifier

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=124751&lokasi=lokal>

Abstrak

Seiring derasnya arus informasi berita elektronik, timbul kebutuhan untuk mengatur informasi tersebut sehingga pengguna dapat mengaksesnya dengan lebih mudah. Akan tetapi jika pengelompokan berita dilakukan secara manual, maka akan memakan waktu yang lama dan mahal. Klasifikasi dokumen secara otomatis sekiranya diperlukan untuk mengurangi biaya dan mempercepat pengaturan informasi. Salah satu metode yang dapat digunakan dalam pengklasifikasian yaitu Naïve Bayes Classifier. Fokus penelitian ini adalah meneliti karakteristik Naïve Bayes Classifier untuk memperoleh kinerja yang optimal dalam proses klasifikasi. Cara yang diterapkan pada penelitian ini yaitu dengan mengujicobakan metode tersebut dengan 3 perlakuan yaitu membandingkan kinerja sistem terhadap stemming maupun non stemming, berbagai proporsi dokumen pembelajaran dan jumlah kategori dalam klasifikasi. Tahapan penelitian dilakukan mulai dari studi pustaka, menerapkan metode Naïve Bayes Classifier dalam pengklasifikasian berita berbahasa Indonesia, melakukan uji coba dan analisa mengenai karakteristik metode ini serta menarik kesimpulan dari hasil analisa. Penelitian dilakukan terhadap 1351 dokumen berita berbahasa Indonesia dari situs www.suarapembaruan.com yang diambil pada bulan Januari 2004 sampai dengan bulan November 2004.

Hasil penelitian menunjukkan bahwa Naive Bayes Classifier merupakan metode yang memiliki tingkat akurasi yang tinggi. Kinerja akan lebih baik jika metode ini diterapkan dengan stemming dibanding tanpa stemming walaupun selisih kinerja keduanya tidak terpaut jauh yaitu sekitar 3,87%. Selain itu, hasil penelitian juga menunjukkan bahwa kinerja metode ini dipengaruhi oleh jumlah dokumen pembelajaran. Semakin banyak dokumen pembelajaran yang digunakan, maka akan semakin tinggi tingkat keakuratan metode ini. Hal ini terbukti dari uji coba kombinasi stemming dengan proporsi dokumen pembelajaran 90% yang mampu mencapai kinerja tertinggi selama penelitian yaitu recall sebesar 93,5%, precision 94,125% dan F-measure 93,81%. Hal menarik yang terjadi adalah akurasi masih tetap relatif tinggi walaupun dokumen pembelajaran secara ekstrim dikurangi menjadi 10%. Hal ini ditunjukkan dengan recall sebesar 89,82%, precision 90,36% dan F-measure 90,1%.

Pada penelitian ini juga mengamati apakah Naïve Bayes Classifier merupakan metode klasifikasi yang stabil. Hal ini diteliti dengan membandingkan kinerja sistem terhadap banyaknya jumlah kategori dalam klasifikasi. Hasilnya ternyata jumlah kategori tidak mempengaruhi kinerja metode ini. NBC merupakan metode yang stabil jika dilihat dari segi kuantitas kategori.