

Klasifikasi topik menggunakan metode naive bayes dan maximum entropy pada artikel media massa dan abstrak tulisan

Dyta Anggraeni, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=123561&lokasi=lokal>

Abstrak

Klasifikasi topik adalah proses pembagian dokumen sesuai dengan topik yang terkandung dari dokumen tersebut. Dalam melakukan klasifikasi topik, pada tugas akhir ini digunakan metode Naïve Bayes dan Maximum Entropi dengan dua jenis data, yaitu artikel media massa dan abstrak tulisan ilmiah dari sistem Lontar. Percobaan ini dilakukan dan dianalisis dari beberapa aspek yaitu metode dan fitur yang digunakan, banyak topik yang digunakan, dan jenis data yang digunakan.

Hasil percobaan yang didapat adalah nilai akurasi tertinggi didapat pada saat menggunakan metode Naïve Bayes dengan informasi fitur frequency-normalized yaitu 95,73%. Selain itu, jumlah token yang semakin banyak digunakan secara umum akan meningkatkan nilai akurasi dan pemakaian abstrak tulisan ilmiah memberikan nilai akurasi yang hampir mirip dengan pemakaian artikel media massa.

<hr>

Topic Classification is a process of categorizing document based on the topic contained in a document. To carry out the topic classification, we use Naïve Bayes and Maximum Entropy towards mass media article and abstracts of scientific papers from Lontar System. Experiments have been done and analyzed regarding several aspects, namely the methods and features, the number of topics, and the data.

In this thesis, we found that Naïve Bayes with frequency-normalized as feature information yield the highest accuracy, 95,73%. Furthermore, as the number of the tokens used increase, the accuracy also increases. Experiments using the abstracts of scientific papers yield similar accuracy to mass media article.