

ABSTRAK

Aldi Sunantyo Ali Murdianto. 2013. : Klasifikasi Berita Berbahasa Inggris Menggunakan Algoritma *K-Nearest Neighbor* (KNN) Berbasis Ontologi

Dosen Pembimbing : Lailil Muflikhah, S.Kom., M.Sc dan Indriati ST,M.Kom

Penelitian ini membahas penerapan algoritma *K-Nearest Neighbor* (KNN) berbasis ontologi untuk mengklasifikasikan teks berita berbahasa Inggris. Konsep klasifikasi teks adalah memasukkan suatu dokumen baru ke dalam suatu kategori pada sekumpulan dokumen. Proses penentuan kategori dari dokumen baru tersebut adalah mencari kemiripan antara dokumen baru dengan sekumpulan dokumen tersebut. Tetapi penentuan kemiripan ini terbatas pada kemiripan antar kata yang memiliki huruf penyusun yang sama. Untuk mengatasi kemiripan antar kata yang berbeda huruf penyusunnya tetapi sama maknanya maka dibutuhkan ontologi atau hubungan semantik antar kata. Sistem yang akan dibangun menggunakan basis ontologi *WordNet* untuk menemukan *term* yang saling bersinonim di dalam dokumen latih dan dokumen uji. *Term* yang saling bersinonim akan digabungkan berikut frekuensi kemunculannya. Algoritma KNN adalah algoritma pembelajaran dimana dokumen uji akan melakukan pelatihan dengan setiap dokumen latih. Dataset yang digunakan adalah *reuters-21578* 90 kategori. Pengujian dari algoritma ini adalah mengetahui nilai *document frequency threshold* dan jumlah “*k*” tetangga yang optimal untuk mengklasifikasikan dokumen. Hasil dari pengujian ini akan dijadikan parameter pengujian selanjutnya yaitu pengaruh jumlah dokumen latih terhadap hasil klasifikasi. Hasil evaluasi sistem menunjukkan nilai *f1-measure* tertinggi adalah 0.949 dengan nilai *threshold*=1 dan nilai “*k*” yang beragam yaitu 12,14,15,17,18,19. Pada pengujian selanjutnya jumlah dokumen latih yang digunakan untuk pengujian berturut-turut adalah 40, 80, 120, 160, 200, dan 400 sementara dokumen uji ditentukan tetap jumlahnya yaitu 20. Pada setiap pengujian tersebut hasil *f1-measure* selalu diatas 0.85 dengan parameter *threshold* dan jumlah “*k*” berdasarkan pada pengujian sebelumnya. Penelitian ini menunjukkan bahwa algoritma *K-Nearest Neighbor* (KNN) berbasis ontologi memiliki kinerja yang baik dalam mengklasifikasikan berita berbahasa Inggris.

Kata kunci : klasifikasi dokumen, KNN, ontologi, *WordNet*

ABSTRACT**Aldi Sunantyo Ali Murdianto. 2013. : Classification of English News Text Using K-Nearest Neighbor (KNN) Based Ontology Algorithm****Advisor: Lailil Muflikhah, S.Kom., M.Sc dan Indriati ST, M.Kom**

This research implements K-Nearest Neighbor (KNN) based ontology algorithm to classify English news text. The principal idea of text classification is we put a document with unknown category into a category from a corpus of text. To determine the category for this unknown category we use learning method to find similarity between this unknown document and each of document in the corpus. This learning method has a weakness where it only can handle the terms which have same morphological form. To overcome this weakness, we can use ontology so we can handle the similarity for terms that have different morphological form but they have similar semantic meaning. The system that will be built is using WordNet database to find similar meaning between term both in corpus and in the unknown document. Term that has similar meaning will be merged including their term frequency. KNN algorithm is learning algorithm where the unknown document will pass learning process with each document in the corpus. Reuters-21578 with 90 categories will be used as dataset. The experiment for this algorithm tests document frequency threshold value and “k” neighbor value that give the best evaluation value for classifying document. This experiment result show that highest f1-measure score is 0.949 with threshold=1 and various “k” neighbor value which are 12, 14, 15, 17, 18, 19. Another experiment tests the influence of number training document with evaluation result. The number of training documents that will be used are 40, 80, 120, 160, 200, and 400 while test document will be constant in 20. This experiment result show that f1-measure value always higher than 0.85 with threshold and “k” neighbor value is based from previous experiment. This research show that K-Nearest Neighbor (KNN) based ontology algorithm has good performance to classify English news text.

Keyword : text classification, KNN, ontology, WordNet.