

ABSTRAK

UPIK atau Unit Pelayanan Informasi dan Keluhan merupakan suatu *platform* milik Pemerintah Kota Yogyakarta yang berfungsi untuk menampung pengaduan berupa saran, informasi, keluhan, atau pertanyaan dari masyarakat Kota Yogyakarta. Pengaduan atau selanjutnya disebut informasi dan keluhan yang masuk kemudian akan dikategorikan ke kategori bidang dan unit kerja yang menangani permasalahan tersebut. Untuk mempermudah mengorganisir informasi dan keluhan, meningkatkan efisiensi waktu *administrator* dalam memilah dan menentukan kategori bidang serta unit kerja, maka perlu dibuat sistem cerdas yang dapat mengklasifikasikan dokumen sesuai tujuan. Informasi dan keluhan memiliki jumlah data sampel perkategori yang tidak seimbang atau dalam klasifikasi teks disebut *imbalanced class*. Informasi dan keluhan juga memiliki ambiguitas kalimat yang memungkinkan suatu informasi dan keluhan dapat dikategorikan ke beberapa kategori bidang dan unit kerja.

Fuzzy K-Nearest Neighbor merupakan algoritma klasifikasi perbaikan dari *K-Nearest Neighbor* dalam penentuan keanggotaan kelas dari suatu data uji. Penggunaan teori *fuzzy* dalam *K-Nearest Neighbor* dapat menangani ambiguitas suatu kalimat. Proses yang dilakukan adalah mengumpulkan data latih dan data uji, melakukan tahap *preprocessing* (diantaranya *case folding, cleansing, tokenization, spelling normalization, stopword removal dan stemming*), pembobotan fitur dengan metode *Term Frequency Invers Document Frequency*, kemudian dilakukan tahap klasifikasi, dan pada tahap akhir dilakukan pengujian menggunakan tabel *confusion matrix multiclass* untuk mencari nilai akurasi, presisi, dan *recall*.

Hasil pengujian menunjukkan nilai akurasi optimum untuk klasifikasi bidang sebesar 96,528% dengan $K=11$ pada persentase data 80(*train*):20(*test*) dan untuk klasifikasi unit kerja sebesar 97,375% dengan $K=9$ pada persentase data 70 (*train*):30(*test*). Sedangkan nilai presisi dan *recall* optimum untuk klasifikasi bidang sebesar 56,597% dengan $K=11$ pada persentase data 80(*train*):20(*test*). Sedangkan untuk klasifikasi unit kerja sebesar 58,052% dengan $K=5$ pada persentase data 90(*train*):10(*test*).

Kata Kunci: *Text Mining, Klasifikasi Teks, Fuzzy K-Nearest Neighbor, Imbalanced Class, Ambiguitas, Confusion Matrix Multiclass*