

INTISARI

KLASIFIKASI JENIS KELAMIN PENULIS MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOUR DAN HARMONY SEARCH

Erick

10/300241/PA/13188

Author profiling adalah suatu proses penentuan profil pengarang dari suatu data tulisan tertentu. Adapun profil dari suatu pengarang adalah kumpulan karakteristik dari pengarang tersebut yang bisa berupa jenis kelamin, usia, kebangsaan ataupun kepribadian.

Banyak algoritma yang dapat digunakan dalam mengimplementasikan *author profiling*. Salah satu algoritma yang dapat digunakan adalah *k-Nearest Neighbour* (*k-NN*). Algoritma ini digunakan untuk mencari *k* buah data yang memiliki kemiripan terbesar dengan data uji yang kemudian data uji tersebut diklasifikasikan ke dalam kelas mayoritas dari *k* buah data yang memiliki kemiripan tertinggi. Jumlah fitur yang tidak sedikit yang digunakan pada penelitian dapat menyebabkan terjadinya permasalahan *curse of dimensionality* pada *k-NN*, sehingga digunakan algoritma *Harmony Search* untuk mengimplementasikan *feature selection* dengan harapan jumlah fitur yang digunakan dapat teroptimalkan.

Berdasarkan pengujian sistem yang telah dilakukan, klasifikasi profil penulis menggunakan algoritma *k-Nearest Neighbour* dan *Harmony Search* mendapatkan akurasi terbaik sebesar 61% dengan recall dan precision yang diperoleh masing-masing sebesar 54% dan 62.79%.

Kata kunci: author profiling, k-NN, HSA

ABSTRACT

AUTHORS' GENDER CLASSIFICATION USING K-NEAREST NEIGHBOUR AND HARMONY SEARCH ALGORITHMS

Erick

10/300241/PA/13188

Author profiling is a task to determine profile of the author of some specific documents. Where as profile of an author may consist of some or all of the author's characteristics such as gender, age, nationality, or personality.

There are a lot of algorithms that can be used in implementing author profiling. One of them is *k*-Nearest Neighbour (*k*-NN). In *k*-NN, the test data is classified to the majority class of the *k* datas that have the highest similarity with the test data. As the numbers of features that are not in a small amount, it may lead to the curse of dimensionality problem in *k*-NN, Harmony Search Algorithm is used in order to implement feature selection so that the number of features can be optimized.

From various testing scheme through the system, authors' profile classification using *k*-Nearest Neighbour and Harmony Search Algorithm yields accuracy with the maximum value at 61% with recall and precision each at the value of 54% and 62.79%.

Keywords: author profiling, *k*-NN, HSA