

Implementasi Algoritma Binary Particle Swarm Optimization (BPSO) dan C4.5 Decision Tree untuk Deteksi Kanker Berdasarkan Klasifikasi Microarray Data

Amalya Citra Pradana¹, Adiwijaya^{2,3}, Annisa Aditsania³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹amalyacitra@students.telkomuniversity.ac.id, ²adiwijaya@telkomuniversity.ac.id,

³aaditsania@telkomuniversity.ac.id

Abstract

Cancer is one of deadly disease in the world. Cancer can be detected by representing the cancer into microarray data with measuring the changes occurred in gene expression level. Cancer detection can be done by doing classification technique for microarray data. One of most algorithm that applied for classification is Decision Tree C4.5. It is a linier method which is easy to interpret and included into the algorithm which has given impact in classification but it is sensitive to noise data. Microarray data has a large features (high dimensional) which is not all features have important information (high noise) and has a small samples and causing the application is difficult and affected the accuracy. Binary Particle Swarm Optimization (BPSO) is one of searching optimization algorithm that could find an optimal feature. Rule in Decision Tree is modelled with discrete value so the data has to be discretized. Discretization is applied using K-Means. System is divided into two schemas such as Information Gain (IG) – C4.5 and BPSO – C4.5. The accuracy based on IG – C4.5 and BPSO – C4.5 schema are 54% and 99%. Feature selection has given impact to classification for avoiding noise data to build the rule accurately. With applying BPSO as feature selection can find the features significantly.

Keywords: microarray data, binary particle swarm optimization, C4.5 decision tree, classification, feature selection, K-Means
