

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

Amalya Citra Pradana¹, Adiwijaya², Annisa Aditsania³

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

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

³aaditsania@telkomuniversity.ac.id

Abstrak

Kanker merupakan salah satu penyakit yang mematikan di dunia. Upaya pendeteksian kanker dapat dilakukan dengan merepresentasikan kanker ke dalam *microarray data* dengan mengukur perubahan yang terjadi pada level ekspresi gen. Deteksi gejala kanker dapat dilakukan dengan teknik *data mining*, yaitu klasifikasi terhadap *microarray data*. Salah satu penerapan algoritma untuk klasifikasi adalah *C4.5 Decision Tree* dimana algoritma tersebut mudah diinterpretasi dan termasuk paling berpengaruh dalam klasifikasi namun memiliki kekurangan yaitu sensitif terhadap data *noise*. *Microarray data* memiliki jumlah *feature* yang sangat besar (*high dimensional*) dimana tidak semua *feature* tersebut memiliki informasi yang penting (*high noise*) dan jumlah sampel yang sedikit sehingga penerapan proses klasifikasi saja menjadi sulit karena dapat mempengaruhi nilai akurasi. *Binary Particle Swarm Optimization* (BPSO) merupakan salah satu algoritma optimasi pencarian untuk mendapatkan fitur yang optimal. Pemodelan *rule* pada *Decision Tree* menggunakan nilai diskrit sehingga data perlu didiskritkan. Diskritisasi dilakukan menggunakan K-Means. Sistem dibagi menjadi dua skema yaitu skema *Information Gain* (IG) – C4.5 dan skema BPSO – C4.5. Akurasi yang diperoleh berdasarkan skema IG-C4.5 dan BPSO-C4.5 berturut-turut adalah 54% dan 99%. Pengaruh seleksi fitur terhadap klasifikasi berperan penting dalam menghindari data *noise* untuk memodelkan *rule* yang akurat. Dengan penerapan BPSO sebagai seleksi fitur mampu mencari fitur yang paling signifikan.

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