

Pengembangan kaidah keterhubungan (associative rules mining) phenotype dan genotype untuk prediksi sekuen gen = The development of associative rules mining of phenotype and genotype for predictive sequence gene / Ranny

Ranny, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20350533&lokasi=lokal>

Abstrak

ABSTRAK

Data genotype merupakan komponen yang penting pada bidang bioinformatik, namun data genotype sulit diperoleh. Selain sulitnya proses yang harus dilakukan untuk memperoleh data genotype, proses tersebut juga hanya dapat dilakukan oleh para ahli yang memiliki kemampuan mengesktrak data genotype. Oleh karena hal tersebut maka pada penelitian ini akan dibangun sebuah sistem yang dapat membentuk rule phenotype-genotype yang digunakan untuk memprediksi genotype berdasarkan rule tersebut. Pembentukan rule akan menggunakan metode association rules mining (ARM). Salah satu algoritma yang menggunakan dasar metode ARM adalah algoritma classification predictive base on association rule (CPAR). Algoritma CPAR menjadi acuan untuk membentuk rule genotypephenotype pada riset ini. Proses prediksi dilakukan dengan menghitung nilai kemiripan antara phenotype pada rule dengan input phenotype yang akan diprediksi genotype-nya. Phenotype dengan nilai batas kemiripan < 0.05 akan merujuk pada rule genotype dan menjadi hasil prediksi. Evaluasi dilakukan dengan menghitung akurasi berdasarkan ground truth. Hasil prediksi mencapai akurasi 48% dengan standar deviasi sebesar 30%.

ABSTRACT

Genotype data is an important component in bioinformatics research; unfortunately it is hard to get the data. An expert is also needed to extract the gene sequence, so that the purpose of this research is to develop a system that can build a genotype-phenotype rules that can predict the gene. This research use association rules mining (ARM) to build the rules. Classification predictive base on association rules (CPAR) is one of ARM algorithm. In this research, CPAR is an algorithm to build the genotype-phenotype rule that can be used to predict the genotype data. The predictive process is based on the similarity between phenotype on the input and the rule. The input phenotype with similarity threshold < 0.05 will refer to the genotype rule. The result will be evaluated by the accuracy based on the ground truth. The average of the accuracy is 48% with standard deviation is around 30%.