

**IMPLEMENTASI DAN ANALISA WAKTU KOMPUTASI PADA
ALGORITMA *RANDOM FOREST* DENGAN *PARALLEL
COMPUTING* DI R**

Nur Azizah (NIM 1001131) azizah.nur@student.upi.edu

ABSTRAK

Random forest merupakan metode untuk membangun model dengan menggabungkan *decision trees* atau pohon keputusan yang dihasilkan dari sampel *bootstrap* dan fitur acak. Permasalahan umum yang sering terjadi pada saat mengimplementasikan *random forest* adalah waktu pemrosesan yang lama karena menggunakan data yang banyak dan membangun model *tree* yang banyak pula untuk membentuk *random trees* karena menggunakan *single processor*. Penelitian ini mengusulkan metode *random forest* dengan *parallel computing* dan diimplementasikan dalam bahasa pemrograman R. Beberapa kasus yang digunakan dalam penelitian ini yaitu *dataset* bunga Iris, kualitas *wine* dan data diagnosa diabetes wanita Pima Indian. Hasil yang diperoleh dari penelitian secara keseluruhan menunjukkan waktu komputasi yang digunakan saat menjalankan *random forest* dengan *parallel computing* lebih singkat dibandingkan dengan saat menjalankan *random forest* biasa yang hanya menggunakan *single processor*.

Kata kunci : *Decision trees, random forest, parallel computing, bahasa pemrograman R*

**IMPLEMENTASI DAN ANALISA WAKTU KOMPUTASI PADA
ALGORITMA *RANDOM FOREST* DENGAN *PARALLEL
COMPUTING* DI R**

Nur Azizah (NIM 1001131) azizah.nur@student.upi.edu

ABSTRACT

Random forests are a set of methods constructing a model by assembling DTs that are generated from bootstrap samples and a randomized features. A common problem that often occurs when implementing random forest is long processing time because it uses a lot of data and build many tree models to form random trees because it uses single processor. This research proposes random forest method with parallel computing and implemented in R programming language. Some of the cases used in this research are Iris flower dataset, wine quality and diabetes diagnosis data of Pima Indian woman. The results obtained from the entire study show that the computational time used when running random forest with parallel computing is shorter than when running a regular random forest using only a single processor.

Keywords : *Decision trees, random forest, parallel computing, bahasa pemrograman R*