

INTISARI

PENGARUH *PARTICLE SWARM OPTIMIZATION* (PSO) TERHADAP *RANDOM FOREST* DAN *EXTREMELY RANDOMIZED TREE (EXTRA TREE)* SEBAGAI SELEKSI FITUR PADA ANALISIS KLASIFIKASI

Oleh

Febina Nur Sabilla Azzahra

19/439210/PA/19033

Beberapa algoritma *tree-based machine learning* sering digunakan di bidang medis untuk memprediksi dan mendiagnosis penyakit tahap awal, seperti algoritma klasifikasi *Extremely Randomized Tree (Extra Tree)* dan *Random Forest*. Namun, kebanyakan data yang tersimpan dalam sistem medis sangat luas dan kompleks, sehingga sering kali data tersebut berisi fitur-fitur yang tidak signifikan yang dapat menyedatkan algoritma klasifikasi dan berdampak negatif pada hasil. Oleh karena itu, perlu dilakukan seleksi fitur yang bertujuan untuk mengeliminasi fitur-fitur yang tidak relevan dari dataset, sehingga dapat meningkatkan akurasi hasil prediksi dengan mereduksi dimensi data. Salah satu metode yang dapat digunakan untuk seleksi fitur, yaitu *Particle Swarm Optimization* (PSO). Pada skripsi ini dilakukan analisis pada data Diagnostik Kanker Payudara Wisconsin dengan menggunakan *Extra Tree* dan *Random Forest*, serta dilakukan seleksi fitur menggunakan metode PSO. Kemudian akan dibandingkan performa kedua model klasifikasi sebelum dan setelah dilakukan seleksi fitur. Dari analisis yang dilakukan diperoleh kesimpulan bahwa metode seleksi fitur PSO mampu meningkatkan performa klasifikasi yang lebih baik untuk kedua model klasifikasi, serta model *Extra Tree* yang dikombinasikan dengan PSO menunjukkan performa klasifikasi terbaik diikuti oleh model *Random Forest* yang dikombinasikan dengan PSO, model *Extra Tree*, dan model *Random Forest*.

Kata kunci : klasifikasi, seleksi fitur, random forest, extremely randomized tree, particle swarm optimization.

ABSTRACT

THE EFFECT OF PARTICLE SWARM OPTIMIZATION (PSO) ON RANDOM FOREST AND EXTREMELY RANDOMIZED TREE (EXTRA TREE) AS FEATURE SELECTION IN CLASSIFICATION ANALYSIS

By

Febina Nur Sabilla Azzahra

19/439210/PA/19033

Several tree-based machine learning algorithms are often used in the medical field to predict and diagnose early stage diseases, such as the Extremely Randomized Tree (Extra Tree) and Random Forest classification algorithms. However, much of the data stored in medical systems is vast and complex, so it often contains insignificant features that can mislead classification algorithms and negatively impact results. Therefore, it is necessary to carry out feature selection which aims to eliminate irrelevant features from the dataset, so that it can increase the accuracy of prediction results by reducing the dimensions of the data. One method that can be used for feature selection is Particle Swarm Optimization (PSO). In this thesis, analysis was carried out on Wisconsin Breast Cancer Diagnostic data using Extra Tree and Random Forest, and feature selection was carried out using the PSO method. Then the performance of the two models will be compared before and after feature selection. From the analysis carried out, it was concluded that the PSO feature selection method was able to improve better classification performance for both classification models, and the Extra Tree model combined with PSO showed the best classification performance followed by the Random Forest model combined with PSO, Extra Tree model, and Random Forest model.

Keywords : classification, feature selection, random forest, extremely randomized tree, particle swarm optimization.