

# Perbaikan *Performance* Algoritma C4.5 AdaBoost Berbasis Optimasi Proses Selektif Fitur

Oleh:

Gede Ajus Setiawan, NIM 1829101001

Program Studi Ilmu Komputer

## ABSTRAK

*Data mining* adalah proses yang mempergunakan teknik statistic dan matematika untuk mengekstraksi dan mengidentifikasi informasi penuh manfaat serta wawasan yang terakit dari aneka database besar. Terdapat berbagai macam teknik yang dikembangkan guna memaknai sebuah data agar dapat ditelaah dalam bentuk pengetahuan. Salah satu metode yang dimaksud adalah metode klasifikasi C4.5. berbagai kasus *binary classification* telah dilakukan uji penyelesaian masalah menggunakan metode C4.5, namun pada literatur yang melakukan uji didapati algoritma C4.5 berbasis *Adaptive Boosting* (AdaBoost) tidak begitu signifikan dalam melakukan peningkatan performasi algoritma C4.5. Sehingga diperlukan suatu penelitian yang menerapkan parameter dan perlakuan yang sama pada dataset yang berbeda untuk mengukur seberapa besar keefektifan AdaBoost dalam meningkatkan performasi algoritma C4.5. Untuk mengetahui tingkat *performance* tersebut, diperlukan studi kasus dari data yang berbeda, ukuran ataupun karakteristik data yang berbeda, dan juga dengan atribut serta beberapa hasil data yang ada pada studi kasus yang berbeda (Novianti & Santosa, 2016). Untuk kasus pada penelitian kali ini, penulis menggunakan 4 (Empat) dataset yang berbeda diantaranya ialah *Airline Passenger Satisfaction*, *IRIS*, *TIC TAC TU*, dan *Water Potability*. Dimana semua dataset tersebut diambil dari Kaggle. Dengan memberikan perlakuan yang sama pada setiap dataset membuat data tersebut lebih terarah dan mudah untuk di analisis serta mempermudah untuk mengetahui seberapa besar efektifitas dari AdaBoost dalam meningkatkan performasi algoritma C4.5 dalam menangani suatu data. Hasil yang diperoleh dari kajian ini berupa model evaluasi dari tahapan masing-masing sampel kasus klasifikasi sebelum dan sesudah mempergunakan algoritma C4.5 berbasis AdaBoost pada setiap dataset yang berbeda dengan perlakuan yang sama. Sehingga didapatkan tahapan dalam *preprocessing* yaitu menerapkan selektif fitur yang sesuai pada kriteria setiap dataset.

Kata kunci : C4.5 Berbasis AdaBoost; selektif fitur dataset; *machine learning*.

***Performance Improvement of the C4.5 AdaBoost Algorithm Based on  
Optimization Feature Selective Process***

**By:**

**Gede Ajus Setiawan, NIM 1829101001**

**Program Studi Ilmu Komputer**

**ABSTRACT**

*Data mining is a process that uses statistical dan mathematical to extract and identify useful information and assembled knowledge from large databases. There are various kinds of techniques developed to make sense of data so that it can be studied in the form of knowledge. One of the methods in question is the C4.5 classification method. Various binary classification cases have been tested for problem-solving using the C4.5 method, but in the literature that did the test, it was found that the C4.5 algorithm based on Adaptive Boosting (AdaBoost) was not so significant in increasing the performance of the C4.5 algorithm. So we need a study that applies the same parameters and treatment to different datasets to measure how effective AdaBoost is in increasing the performance of the C4.5 algorithm. To find out the level of performance, case studies are needed from different data, different data sizes or characteristics, and also with attributes and some of the results of the data in different case studies (Novianti & Santosa, 2016). For the case in this study, the authors used 4 (four) different datasheets including Airline Passenger Satisfaction, IRIS, TIC TAC TU, and Water\_Potability. Where all the datasets are taken from Kaggle. By giving the same treatment to each dataset, makes the data more focused and easier to analyze and makes it easier to find out how effective AdaBoost is in improving the performance of the C4.5 algorithm in handling data. The results obtained from this study are in the form of an evaluation model of the stages of each sample case of classification before and after using the AdaBoost-based C4.5 algorithm on each different dataset with the same treatment. So that the stages in preprocessing are obtained, namely applying selective features that match the criteria for each dataset.*

*Keywords : C4.5 Based on AdaBoost; selective dataset features; machine learning.*