

ABSTRAK

Oswaldo Figo (01112180010)

PERBANDINGAN PERFORMA *BAGGING* DAN *ADABOOST* DENGAN BASIS *LOGISTIC REGRESSION* UNTUK MENGLASIFIKASI DATA *MULTI-CLASS*

Skripsi, Fakultas Sains dan Teknologi (2022).

(xiv + 65 halaman; 16 tabel; 42 gambar)

Salah satu teknik yang paling populer digunakan untuk meningkatkan performa sebuah algoritma *Machine Learning* adalah menggunakan *Ensemble Learning*. Ide dari teknik ini adalah dengan menggabungkan beberapa algoritma *Machine Learning* atau yang biasa disebut sebagai *base learners*. Penelitian ini menggunakan algoritma atau metode *Logistic Regression* sebagai *base learners* untuk membentuk model *Ensemble Learning*. Tujuan utama dari penelitian ini adalah membandingkan performa (nilai evaluasi dan waktu) dua algoritma *Ensemble Learning* yaitu metode *Bootstrap Aggregating (Bagging)* dan metode *Adaptive Boosting (AdaBoost)*. Penelitian menggunakan sebelas *dataset* dengan klasifikasi *multi-class* yang independen terhadap karakteristik (proporsi data, jumlah data, dan masalah) serta jumlah kelas variabel target berbeda. Hasil penelitian menunjukkan bahwa model yang dibentuk oleh metode *Bagging* cenderung menunjukkan performa nilai yang lebih baik dari metode *AdaBoost* pada metrik evaluasi jika hanya dilihat berdasarkan per *dataset* yang digunakan, dengan rata-rata nilai evaluasi sebesar 72,21% dan 61% untuk *Bagging* serta 66,25% dan 53,7% untuk *AdaBoost* dengan metrik akurasi dan F1 secara berurutan dalam pembentukan model. Namun, hasil uji rataan untuk kedua metode menjelaskan bahwa tidak ada metode yang lebih baik dari metode lainnya atau dapat dikatakan bahwa performa metode *Bagging* sama dengan performa metode *AdaBoost*. Hasil uji rataan untuk waktu pembentukan kedua model juga memberikan hasil yang tidak signifikan atau dapat dikatakan bahwa lama waktu komputasi untuk membentuk model *Bagging* dan model *AdaBoost* sama.

Kata kunci : *Ensemble Learning, Bagging, Boosting, AdaBoost, Logistic Regression multi-class classification, Ensemble Logistic Regression.*

Referensi : 21 (1996-2021)

ABSTRACT

Oswaldo Figo (01112180010)

COMPARISON OF BAGGING AND ADABOOST PERFORMANCE WITH LOGISTIC REGRESSION BASIS TO CLASSIFY MULTI-CLASS DATA

Thesis, Faculty of Science and Technology (2022).

(xiv + 65 pages, 16 tables, 42 figures)

One of the most popular techniques used to improve the performance of a Machine Learning algorithm is using Ensemble Learning. The idea of this technique is to collect several Machine Learning algorithms or commonly referred to as base learners, which means an Ensemble Learning is a combination of several base learners. This study uses the Logistic Regression algorithm or method as base learners to form an Ensemble Learning model. The main objective of this research is to compare the performance (evaluation score and time) of two Ensemble Learning algorithms, namely the Bootstrap Aggregating (Bagging) method and the Adaptive Boosting (AdaBoost) method. This study uses eleven datasets with multi-class classifications that are independent of the characteristics (data proportion, number of data, and problems) also the number of different classes of target variables. The results show that the model formed by the Bagging method tends to show better value performance than the AdaBoost method based on evaluation metrics if it is only seen based on the dataset used, with an average evaluation value of 72.21% and 61% for Bagging and 66.25% and 53.7% for AdaBoost with accuracy metrics and F1 respectively in model building. However, the mean test results for both methods explain that neither method is better than the other or it can be said that the performance of the Bagging method is the same as the performance of the AdaBoost method. The mean test results for the formation time of the two models also give insignificant results or it can be said that the computation time for the Bagging model and the AdaBoost model is the same.

Keywords : *Ensemble Learning, Bagging, Boosting, AdaBoost, Logistic Regression multi-class classification, Ensemble Logistic Regression.*

Reference : 21 (1996-2021)