

INTISARI

PENINGKATAN PERFORMA KLASIFIKASI ALGORITMA C4.5 MENGUNAKAN *SPLIT FEATURE REDUCTION* DAN *BOOTSTRAP AGGREGATING (BAGGING)*

Oleh
Hafizhah Farah Azhar
18/424288/PA/18393

Dalam machine learning, salah satu algoritma klasifikasi yang tergolong dalam kategori supervised learning adalah algoritma C4.5. Pada dataset dengan jumlah fitur yang besar, dimungkinkan adanya fitur-fitur yang tidak relevan dengan proses klasifikasi yang menyebabkan performa model klasifikasi menjadi tidak optimal. *Split feature reduction* adalah salah satu metode pemilihan fitur yang bekerja berdasarkan pemeringkatan pada nilai gain yang kemudian membagi data menjadi beberapa skema split yang mungkin terbentuk. Dibandingkan dengan metode klasifikasi lain, model pohon keputusan termasuk dalam metode yang tidak stabil dikarenakan perubahan kecil dalam data training dapat menyebabkan perubahan signifikan pada model klasifikasi. Selain itu, terdapat risiko terjadinya *overfitting* pada data dengan variansi besar. Oleh karena itu, diterapkan metode *bagging* untuk meningkatkan stabilitas dan performa klasifikasi. Dari hasil analisis, diketahui bahwa *bagging* meningkatkan performa klasifikasi yakni akurasi, presisi, sensitivitas, spesifisitas, dan skor F-1, serta adanya pembagian data menjadi beberapa *split* mampu memilih kombinasi fitur mana yang menghasilkan performa terbaik.

Kata Kunci: Klasifikasi, C4.5, *Bootstrap Aggregating (Bagging)*, *Split Feature Reduction*.

ABSTRACT

THE IMPROVEMENT OF CLASIFICATION PERFORMANCE IN C4.5 ALGORITHM USING SPLIT FEATURE REDUCTION AND BOOTSTRAP AGGREGATING (BAGGING) METHOD)

By

Hafizhah Farah Azhar

18/424288/PA/18393

In machine learning, one of the classification algorithms belonging to the supervised learning category is the C4.5 algorithm. In datasets with a large number of features, it is possible that there are features that are not relevant to the classification process which causes the performance of the classification model to be not optimal. Split feature reduction is a feature selection method that works based on ranking on the gain value which then divides the data into several possible split schemes. Compared to other classification methods, the decision tree model is an unstable method because small changes in the training data can cause significant changes to the classification model. In addition, there is a risk of overfitting on data with large variances. Therefore, the bagging method is applied to improve the stability and classification performance. From the results of the analysis, it is known that bagging improves classification performance, namely accuracy, precision, sensitivity, specificity, and F-1 score, as well as the division of data into several splits to be able to choose which combination of features produces the best performance.

Keywords: Classification, C4.5, Bootstrap Aggregating (Bagging), Split Feature Reduction.