

## **ABSTRACT**

*Classification using supervised learning in intelligence information or evaluation systems is needed to accelerate process and increase accuracy of human resource decision-making in the organization. However, the ability of this supervised learning depend on high-quality data. This study proposes feature selection using RFE to improve quality data by eliminating irrelevant attributes and identify its impact on the classification ability of one of the supervised learning algorithms, namely Random Forest. Based on experiments on an open dataset from Kaggle called the Human Resource Dataset, application of RFE with the right parameters in this study succeeded in eliminating 16 (64%) irrelevant attributes and increasing 2.6% the average AUC-ROC Random Forest. In addition, the (84.32%) average AUC-ROC of Random Forest with RFE feature selection in this study was higher than Naïve Bayes (76.49%), Logistic Regression (63.40%), Support Vector Machine (56.56%) ) and K-Nearest Neighbors (54.04%).*

**Keywords:** *Employee, Performance, Machine Learning, Random Forest, RFE.*

## ABSTRAKSI

Klasifikasi menggunakan *supervised learning* dibutuhkan pada sistem informasi atau evaluasi cerdas untuk mempercepat proses dan meningkatkan akurasi pengambilan keputusan terkait sumber daya manusia di organisasi. Namun, kemampuan dari *supervised learning* ini relatif bergantung pada kualitas data. Studi ini mengusulkan seleksi fitur menggunakan *Recursive Feature Elimination (RFE)* untuk meningkatkan kualitas data dengan mengeliminasi atribut-atribut yang tidak relevan dan mengidentifikasi dampaknya terhadap kemampuan klasifikasi salah satu algoritma *supervised learning* yaitu *Random Forest*. Berdasarkan hasil eksperimen pada dataset terbuka dari *Kaggle* bernama *Human Resource Dataset*, penerapan *RFE* dengan parameter yang tepat pada studi ini berhasil mengeliminasi 16 (64%) atribut yang tidak relevan dan meningkatkan 2.6% rata-rata *AUC-ROC Random Forest*. Adapun 84.32% rata-rata *AUC-ROC* hasil evaluasi *Random Forest* setelah penerapan *RFE* pada studi ini, lebih tinggi dibandingkan dengan *Naïve Bayes* (76.49%), *Logistic Regression* (63.40%), *Support Vector Machine* (56.56%) dan *K-Nearest Neighbors* (54.04%).

**Kata kunci:** Karyawan, Kinerja, *Supervised Learning*, Seleksi Fitur, *RFE*.

