

ABSTRACT

Employee turnover is a strategic challenge that can significantly affect organizational stability, productivity, and operational sustainability. The ability to accurately predict potential employee departures is therefore essential for organizations to formulate more effective retention strategies. However, the complex and imbalanced nature of turnover data makes the prediction process more challenging. This study aims to compare the performance of the Random Forest and XGBoost machine learning algorithms and to evaluate the effectiveness of a Stacking Ensemble approach that combines both algorithms and is optimized using Bayesian Optimization to enhance predictive performance. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is applied, while a Multi-Layer Perceptron (MLP) is employed as the meta-learner in the stacking model. Based on the evaluation results, the Stacking model with an MLP meta-learner optimized through Bayesian Optimization achieves the best performance, outperforming the individual Random Forest and XGBoost models. This model attains the highest performance with an accuracy of 0.9405, precision of 0.9408, recall of 0.9406, F1-score of 0.9409, and AUC of 0.9789, demonstrating excellent capability in identifying high-risk employees under imbalanced data conditions. This study is expected to serve as a reference for the development of machine learning-based employee turnover prediction strategies and to provide insights for HR practitioners in supporting decision-making related to employee retention management.

Keywords: *Employee Turnover, Random Forest, XGBoost, Stacking Ensemble, MLP, SMOTE, Bayesian Optimization, Machine Learning.*

ABSTRAK

Turnover karyawan merupakan tantangan strategis yang dapat berdampak pada stabilitas, produktivitas, serta keberlanjutan operasional perusahaan. Kemampuan untuk memprediksi potensi keluarnya karyawan secara akurat menjadi penting agar organisasi dapat merumuskan strategi retensi yang lebih efektif. Namun, karakteristik data *turnover* yang kompleks dan tidak seimbang membuat proses prediksi menjadi lebih menantang. Penelitian ini bertujuan untuk membandingkan performa algoritma *Machine Learning Random Forest* dan *XGBoost*, serta mengevaluasi efektivitas pendekatan *Stacking Ensemble* yang mengombinasikan kedua algoritma tersebut dan dioptimalkan menggunakan *Bayesian Optimization* sebagai upaya peningkatan performa prediksi. Untuk menangani ketidakseimbangan kelas digunakan teknik SMOTE, sementara *Multi Layer Perceptron* (MLP) digunakan sebagai *meta-learner* pada *model stacking*. Berdasarkan hasil evaluasi, model *Stacking* dengan *meta-learner* MLP yang dioptimalkan menggunakan *Bayesian Optimization* menghasilkan performa terbaik, mengungguli model individual *Random Forest* maupun *XGBoost*. Model ini memperoleh performa tertinggi dengan *Accuracy* 0.9405, *Precision* 0.9408, *Recall* 0.9406, *F1-Score* 0.9409 dan AUC 0.9789, menunjukkan kemampuan yang sangat baik dalam mengidentifikasi karyawan berisiko tinggi pada kondisi data yang tidak seimbang. Penelitian ini diharapkan dapat menjadi rujukan bagi pengembangan strategi prediksi *turnover* berbasis machine learning, sekaligus memberikan wawasan bagi praktisi SDM dalam mendukung pengambilan keputusan terkait manajemen retensi karyawan.

Kata Kunci: *Turnover Karyawan, Random Forest, XGBoost, Stacking Ensemble, MLP, SMOTE, Bayesian Optimization, Machine Learning.*