

ABSTRACT

Unemployment is a complex issue in developing countries like Indonesia, impacting economic growth and social welfare. The Open Unemployment Rate (OER) serves as a key indicator to assess regional labor market conditions. This study aims to implement the Random Forest algorithm to build a model capable of classifying the Open Unemployment Rate across Indonesian provinces into low, medium, or high categories, and to identify the most influential factors in determining OER classification. Secondary data from the Central Statistics Agency (BPS) for 2015-2024, comprising 340 observations, were processed through several stages. These stages included pre-processing, handling imbalanced data using the Synthetic Minority Oversampling Technique (SMOTE), which increased the dataset to 701 observations, and hyperparameter tuning via Grid Search with 5-fold cross-validation to obtain the best model parameters. The model was evaluated using accuracy, precision, and recall metrics. The results indicate that the Random Forest model, with a 90% training data and 10% test data proportion, yielded the best performance, achieving an accuracy of 92.75%, precision of 92.8%, and recall of 92.7%. Feature importance analysis revealed that the percentage of the poor population was the most influential variable in OER classification, contributing 19.64%.

Keywords: Open Unemployment Rate, Classification, Random Forest, Feature Importance