

International Journal of Population Data Science

Journal Website: www.ijpds.org



The (Un)Predictability of Early (Un)Employment: A Machine Learning Approach

Sanni Kuikka¹

¹Department of Sociology, Demography Unit, Stockholm University, Stockholm, Sweden

This study integrates scholarship on determinants of labor market transitions with the flexible modeling strategies available through machine learning. Leveraging population registers from Finland, this study analyzes how well we can predict (un)employment and for whom we can (not) predict. The study finds that the predictability of long-term (un)employment improves when using tree-based nonparametric models compared to logistic regression but that predictability varies substantially across outcome groups. None of the models predict very accurately for the unemployed net

of baseline likelihood to be unemployed—a group that is well researched and often of interest when designing policies and interventions. Overall accuracy is driven by the employed while incorrectly predicting most of the unemployed. Additionally, the outcomes for individuals with low parental education are overall more difficult to predict than the outcomes for individuals with mid/high parental education, whereas predicting unemployment slightly improves among the low parental education group.

