UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
UNECE Machine Learning for Official Statistics Workshop 2023
5-7 June 2023, Geneva Switzerland
04 April 2023

Progression patterns in the Swiss social security system based on Machine Learning: methods for evaluating quality and model drift

Athanassia Chalimourda, Michael Leuenberger, Brandon Qorri Gonzalez, Luzius von Gunten (Swiss Federal Statistics Office)
athanassia.chalimourda@bfs.admin.ch

Abstract

As part of the innovation project "Machine Learning – Social Security" (ML_SoSi) of the Swiss Federal Statistical Office (SFSO), we studied progressions of new unemployment insurance beneficiaries in the Swiss social security system. These progressions consist of the monthly states in the social security system of an insured person over a period of four years. Hierarchical clustering was used to identify an initial categorization of progression patterns based on a cohort of beneficiaries entering the social security system in 2010. However, these patterns may change over time, due to individual factors, like changes in the beneficiary population in terms of migration background or family composition, as well as to contextual factors, like changes in laws or economic shocks. Starting with the patterns found in the cohort of beneficiaries in 2010 as a reference, we investigated how this reference changes over time. To this end, we transferred the reference to future cohorts using random forest models for prediction. We present results on how the reference patterns based on the cohort of insured persons in 2010 apply to future cohorts, using the cohorts of 2011 and 2015 as examples. Furthermore, we present methodological approaches that help quantify the evolution of the reference patterns in time and answer the principal question when the reference must be updated due to a possible model drift. In doing so, we rely on both internal measures of cluster validation, which characterize the quality of a clustering, and external measures, which compare the reference with alternative patterns of later cohorts of insured persons. Criteria regarding the internal and external measures that result in an actualization of the reference patterns must thus be defined. The transition to a new categorization poses additional challenges. For example, the assignment of existing reference patterns to patterns of a new categorization is not always straightforward. The resulting knowledge gain can in turn be used in establishing new reference patterns. The aspects mentioned above are explained, discussed and potential future developments are shown.