
Organisational Aspects of Implementing ML Based Data Editing in Statistical Production

Steffen Moritz & Sirello, Olivier Olivier.Sirello@bis.org (Destatis, Germany)

Steffen.Moritz@destatis.de

Abstract

The Data Editing Task Team of the UNECE Applying Data Science and Modern Methods Group (ADSaMM) explored during 2023 organizational aspects of implementing machine learning (ML) based data editing in statistical production. Despite the potential identified in the 2019-2020 High-Level Group for the Modernisation of Official Statistics (HLG-MOS) project, the adoption of ML methods for editing and imputation by statistical agencies has remained limited. Therefore, the task team chose to focus on non-methodological barriers that might hinder ML adoption. Utilizing a structured approach, the task team, that involved multiple countries, developed and refined a template for collecting and analyzing use cases from various statistical agencies. This template included detailed descriptions of the methods, challenges and lessons learned of the projects. Overall, seven completed use case templates were supplied. The analysis revealed several critical themes: motivations for adopting ML solutions, challenges in obtaining labeled data, the need for effective collaboration between business, methodological, and IT teams, the importance of domain expertise, the complexity and opacity of ML models, and issues related to IT infrastructure. The final report of the task team offers practical guidelines and recommendations to address these barriers, supported by in-depth reflections and supplementary appendices. This talk gives a short overview of the findings of the task team. By providing these insights, we aim to promote the integration of ML methods in data editing processes, thereby improving the efficiency and quality of official statistical production.