

Virtual UNECE Workshop on Statistical Data Editing

31 August – 4 September 2020

ML for E&I and new and emerging methods

Discussants: Darren Gray (Statistics
Canada), Daniel Kilchmann (SFSO)

Q&A

1. (Dumpert) Did different countries use the same algorithms in different settings in the UNECE HLG-MOS ML for E&I project?
2. (Rocci) Did you test your methods based on the target variable indicating which units have been changed by interactive treatment?
3. (Sthamer) Consistency between total income and income components?

Discussion

1. (general) Is it an issue to keep ML (supervised learning) up-to-date as data used for learning get outdated? Any general procedure?

Discussion

2. (Dumpert) Two countries publishing the same statistics have not always exactly the same data (features) at hand. Could then analysis of the feature importance be easily used to assess the potential loss in prediction accuracy?
3. (Sthamer) Any outlier detection needed? And is the skewed distribution of income of any concern?

Discussion

4. (Rocci) Did you look at the feature importance? If yes, was the feature of selective editing important?
5. (Perkaskaya, Zhang) How to choose the features between the first and second phase (how to determine the features to assess heterogeneous variance/mean in advance)?
6. (Perkaskaya, Zhang) Is it an issue of using 1st phase results (of any method) as features for the 2nd phase?