# On the use of machine learning methods for the treatment of unit nonresponse in surveys

*Marco J.H. Puts, Piet J.H. Daas*[1]

## Abstract

*The paper discusses the important difference between techniques and methods and how this is particularly important when machine learning techniques are applied within the context of official statistics. This is illustrated by discussing four examples of machine learning work performed by the authors which exemplify the importance of a methodological sound approach when including machine learning techniques. The need for training data that represents the target population studied illustrates this without a doubt. This is essential to obtain models that can be applied to the unseen part of the population, particularly when found training data is used.*

**Keywords**: Data science, Methodology, Official Statistics, Representativity.

---

[1]  Statistics Netherlands, m.puts@cbs.nl, pjh.daas@cbs.nl