

**A new generic method to improve machine learning applications in official statistics**

Mr. Kevin Kloos  
Intern Methodologist  
Statistics Netherlands  
(CBS)  
<https://www.cbs.nl>

**Abstract**

The use of machine learning algorithms at national statistical institutes has increased significantly over the past few years. Applications range from new imputation schemes to new statistical output based entirely on machine learning. The results are often very promising. However, recent studies have shown that the use of machine learning in official statistics always introduces a bias, known as misclassification bias. That type of bias does not occur in traditional applications of machine learning and therefore it has received little attention in the academic literature. In earlier work, we have collected existing methods that are able to correct misclassification bias. We have compared their statistical properties, including bias, variance and mean squared error. In this paper, we present a new generic method to correct misclassification bias and we derive its statistical properties. Moreover, we show numerically that it has a lower mean squared error than the existing alternatives in a wide variety of settings. We believe that our new method may improve machine learning applications in official statistics and we aspire that our work will stimulate further methodological research in this area.