

Title: Evaluating Synthetic Data Generation Methods: A Comparative Analysis of CART, Bayesian Networks, and GANs

Authors: Jonathan Latner, Jörg Drechsler, and Marcel Neuenhoffer

Contact email: jonathan.latner@iab.de

Abstract:

In this study, we compare and contrast various methods for generating synthetic data from simulation data to evaluate their strengths and weaknesses. Methods include classification and regression trees (CART), Bayesian network models, and generative adversarial networks (GANs). A comprehensive framework is developed to isolate the core mechanisms each method employs to generate synthetic data, facilitating a nuanced understanding of their operational principles. Key performance metrics, such as fidelity to original data, computational efficiency, and robustness to variations in input parameters, are utilized to evaluate each method. The findings contribute to the broader discourse on synthetic data utility, offering guidance for practitioners and researchers seeking to use synthetic data for enhanced data privacy, resource efficiency, and analytical accuracy.