Data Comparison Tool

Ayush Mishra, Ansh Tyagi, Siddharth Jha, Abhishek Kumar

Department of CSE and ML

Raj Kumar Goel Institute of Technology, Ghaziabad, India

Abstract: *In recent years, synthetic data has become increasingly popular due to its wide range of applications. However, evaluating the usefulness and privacy of synthetic data remains a complex task, as existing assessment methods vary based on the type of data, the intended use case, and the evaluation goal. Currently, there is no universal method for measuring utility and privacy together. To address this, we have created a comprehensive tool that combines 24 modern evaluation techniques into a single executable program. This tool generates reports that compare datasets—synthetic or real—by highlighting similarities and identifying potential privacy risks. The included methods span from visual and statistical utility assessments to privacy risk detection, making it easier for researchers and industry professionals to understand and improve their synthetic datasets. Our goal is to offer a solid foundation for developing more general and reliable evaluation methods in the future.*

Keywords: Synthetic data, data utility, data privacy, dataset comparison, evaluation metrics