1. The intermediate step that uses clustering and diffusion will serve generating data for training deep learning models. This tackles the “not enough samples for training deep learning model” problem. Now based on how much ever data is generated, will be used in the next step of synthetic data generation. Now DeepMTD will be applied even for small datasets and generates synthetic data that closely resembles original data
2. Support of differential privacy
3. Inherently upsample or downsample the synthetic data generated from the original dataset with sensitive information for easy of sharing
4. A larger data can be downsampled with synthetic data for sharing or a small sample can be upsampled with DP support for ease of sharing

Bridge gap between GBDT & DL – Apply the framework for any dataset size and generate artificial data suitable either for deep learning or other machine learning algorithms

Reduce mode collapse

For Loss: Use cycle consistency loss, MMD and discriminator loss

Unlike other VAEs, the original data is not directly used. Pseudo fake samples generated from MTD are the input. Thus differential privacy is achieved