Reproducing the Experiments

1 Datasets

The datasets are stored in the data folder and include:

1. Adult Income

2. Travel Customer

3. Credit Default

4. Heloc

2

The models used for synthetic data generation are:

1. T-VAE

2. CT-GAN

3. ComGen

3 Dependencies

The following libraries should be installed:

1. SDV: pip install sdv

2. Numpy: pip install numpy

3. Pandas: pip install pandas

4. Scikit-learn: pip install scikit-learn

5. Lime: pip install lime

4 Procedure: Baseline Models

1. Use the data_split.ipynb to first read the datasets and then make train test split based on different random states.

2. Use baselines.ipynb to generate synthetic samples using the baseline models. First, read the datasets, generate synthetic samples.

3. Perform evaluation of synthetic samples.

5 Procedure: ComGem

1. Use ComGen.ipynb to generate synthetic samples using ComGen approach

2. First, read the datasets, and generate either differentially private or non private models for each datasets.

3. Discretize the latent space.

4. Generate t-way test using the ACTS tool. ACTS tool is provided in the ACTS 3.2 folder, with instructions on how to use it.

5. Use the t-way samples to generate synthetic saples

6. Perform evaluation of synthetic samples.