# German Dataset

|  |  |
| --- | --- |
| **Data Attributes** | |
| # of data points | 1000 |
| Public Label | 0 = 1  1 = 2 |
| Sensitive Label |  |
| # of numerical and label features | (1000, 13) |
| # of one hot encoding features | (1000, 50) |
| Clean data shape | (1000, 63) |
| Public label statistics | 30.0 percent is 1 |
| Sensitive label statistics | 69.0 percent is 1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | One-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | representation\_size = 100 | | |
| batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] |
| **Test Acc.** | **Pred.** | 0.762 |  | 0.768 |  | 0.759 |
| **Disc.** |  | 0.621 |  | 0. 626 | 0.411 (0.32, 0.67) |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | Three-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 16  epochs = 20  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | batch\_size = 16  epochs = 20  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | representation\_size = 64 | | |
| batch\_size = 16  epochs = 20  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | batch\_size = 16  epochs = 10  learning\_rate = 0.001  hidden\_size = 60  gamma = 1.0  weight = 0 | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] |
| **Test Acc.** | **Pred.** | 0.75 |  | 0.732 |  | 0.718 |
| **Disc.** |  | 0.652 |  | 0. 626 | 0.616 |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | One-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | representation\_size = 100 | | |
| batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] |
| **Test Acc.** | **Pred.** | 0.762 |  | 0.768 |  | 0.759 |
| **Disc.** |  | 0.621 |  | 0. 626 | 0.411 (0.32, 0.67) |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | One-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | representation\_size = 100 | | |
| batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] | batch\_size = 60  epochs = 25  learning\_rate = 0.01  hidden\_size = 100  gamma = 1.0  weight = [0.3,0.7] |
| **Test Acc.** | **Pred.** | 0.762 |  | 0.768 |  | 0.759 |
| **Disc.** |  | 0.621 |  | 0. 626 | 0.411 (0.32, 0.67) |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | |  |  |  | | |
| **Hyper param.** | |  |  |  | | |
|  |  |  |
| **Test Acc.** | **Pred.** |  |  |  |  |  |
| **Disc.** |  |  |  |  |  |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | |  |  |  | | |
| **Hyper param.** | |  |  |  | | |
|  |  |  |
| **Test Acc.** | **Pred.** |  |  |  |  |  |
| **Disc.** |  |  |  |  |  |

# Adult Dataset

|  |  |
| --- | --- |
| **Data Attributes** | |
| # of data points | 32561 |
| Public Label (index = 14) | 0 = <=50K  1 = >50K |
| Sensitive Label (index = 9) | 0 = Male  1 = Female |
| # of numerical and label features | (1000, 6) |
| # of one hot encoding features | (1000, 100) |
| Clean data shape | (1000, 106) |
| Public label statistics | 24.08 percent is 1 |
| Sensitive label statistics | 33.08 percent is 1 |

Biased Dataset

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | One-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | representation\_size = 100 | | |
| batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = -1.0  weight = 0 |
| **Test Acc.** | **Pred.** | 0.845 |  | 0.850 |  | 0.842 |
| **Disc.** |  | 0.829 |  | 0.828 | 0.44 (0.32, 0.67) |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | Three-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 16  epochs = 1  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | batch\_size = 16  epochs = 1  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | representation\_size = 64 | | |
| batch\_size = 16  epochs = 1  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | batch\_size = 16  epochs = 1  learning\_rate = 0.001  hidden\_size = 64  gamma = 1.0  weight = 0 | batch\_size = 16  epochs = 2  learning\_rate = 0.01  hidden\_size = 64  gamma = -1.0  weight = [0.3,0.7] |
| **Test Acc.** | **Pred.** | 0.821 |  | 0.845 |  | 0.846 |
| **Disc.** |  | 0.827 |  | 0.852 | 0.5 |

For NMT, the prediction is either all 1 or 0.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | | One-layer NN | Three-layer NN | One-Layer NN | | |
| **Hyper param.** | | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | representation\_size = 100 | | |
| batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = 1.0  weight = 0 | batch\_size = 500  epochs = 20  learning\_rate = 0.001  hidden\_size = 100  gamma = -1.0  weight = 0 |
| **Test Acc.** | **Pred.** |  |  |  |  | 0.849 |
| **Disc.** |  |  |  |  | 0.44 |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | |  |  |  | | |
| **Hyper param.** | |  |  |  | | |
|  |  |  |
| **Test Acc.** | **Pred.** |  |  |  |  |  |
| **Disc.** |  |  |  |  |  |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | |  |  |  | | |
| **Hyper param.** | |  |  |  | | |
|  |  |  |
| **Test Acc.** | **Pred.** |  |  |  |  |  |
| **Disc.** |  |  |  |  |  |
| **Represent.** | | **X** | | **Encoder** | | |
|  | | **Predictor** | **Discriminator** | **Predictor** | **Discriminator** | **NMT** |
| **Architecture** | |  |  |  | | |
| **Hyper param.** | |  |  |  | | |
|  |  |  |
| **Test Acc.** | **Pred.** |  |  |  |  |  |
| **Disc.** |  |  |  |  |  |

# Experiment Note

|  |  |
| --- | --- |
| Date |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |