Supplementary Information: Navigating Fairness and Privacy

ANONYMOUS AUTHOR(S)

ACM Reference Format:
Anonymous Author(s). 2024. Supplementary Information: Navigating Fairness and Privacy. 1, 1 (September 2024), 26 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

This supplementary material provides additional details and results that complement the findings discussed in the main paper. We include further explanations of the methods, extra figures, and tables for clarity.

2 Mathematical definition of the used metrics

2.1 privacy metrics

- **Jensen-Shannon divergence (JSD):** The Jensen–Shannon divergence (JSD) is a symmetrized and smoothed version of the Kullback–Leibler divergence. It is defined as [3]:

$$\mathrm{JSD}(P, Q) = \frac{1}{2} \left(D_{\mathrm{KL}}(P \parallel M) + D_{\mathrm{KL}}(Q \parallel M) \right)$$

  Where:
  - $D_{\mathrm{KL}}(P \parallel M)$ represents the Kullback-Leibler divergence from distribution P to M.
  - P and Q are the two probability distributions being compared.
  - M is the average distribution, defined as $M = \frac{P+Q}{2}$.
- **Wasserstein distance (WD):** The Wasserstein distance (WD) is defined as [2]:

$$\mathrm{WD}(P, Q) = \int_{\Omega} d(x, y)\, \gamma(x, y)\, d\gamma(x, y)$$

  Where:
  - $\gamma(x, y)$ is the optimal transport plan.
  - $d(x, y)$ is the distance between two points x and y in the probability space.
- **Membership Inference Attack (MIA):** We used the DOMIAS model, a density-based Membership Inference Attack (MIA) model proposed in [4]. This model detects local overfitting in synthetic data generators by comparing the densities of real and synthetic data distributions to infer whether a specific real sample was part of the training data. The membership score is computed as:

$$S(x) = \frac{p_{\mathrm{gen}}(x)}{p_{\mathrm{real}}(x)}$$

  Where:

Manuscript submitted to ACM

- $p_{\text{gen}}(x)$ is the density of the synthetic data.
- $p_{\text{real}}(x)$ is the density of the real data.

A higher score $S(x)$ suggests a greater likelihood that the sample x was used during training, especially in overfitted regions where the synthetic data mimics the real data closely.

- **K-anonymization:** K-anonymity is a property of a dataset that indicates the re-identifiability of its records. A dataset is k-anonymous if quasi-identifiers for each person in the dataset are identical to at least $k-1$ other people in the dataset [1].

3 Results for Same Train, Same Test

3.1 Results Pertaining to RQ1

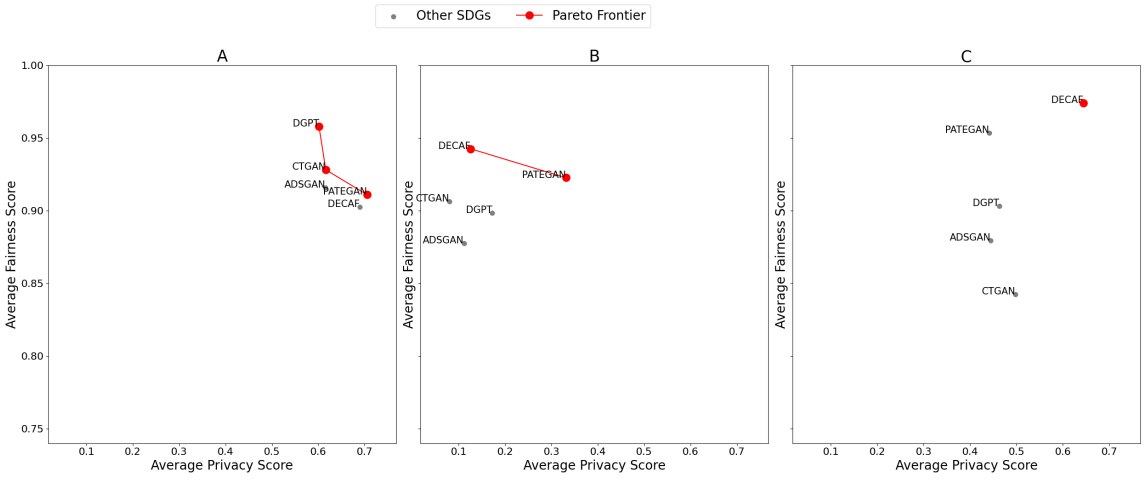

Fig. 1. Pareto frontier illustrating the trade-off between fairness and privacy. Red-labeled points represent optimal solutions, balancing both objectives, while gray points are suboptimal. The frontier highlights the trade-offs between improving fairness and preserving privacy.

Table 1. Fairness and Privacy for all Datasets. **Boldened** scores are the highest and those in red are the lowest.

	Dataset A		Dataset B		Dataset C	
Data	Fairness	Privacy	Fairness	Privacy	Fairness	Privacy
Real Data	0.95	-	0.91	-	0.89	-
ADSGAN	0.92	0.62	0.88	0.11	0.88	0.44
CTGAN	0.93	0.62	0.91	0.08	0.84	0.50
DECAF	0.90	0.69	**0.94**	0.13	**0.97**	**0.64**
DGPT	**0.96**	0.60	0.90	0.17	0.90	0.46
PATEGAN	0.91	**0.71**	0.92	**0.33**	0.95	0.44

3.2 Results Pertaining to RQ2

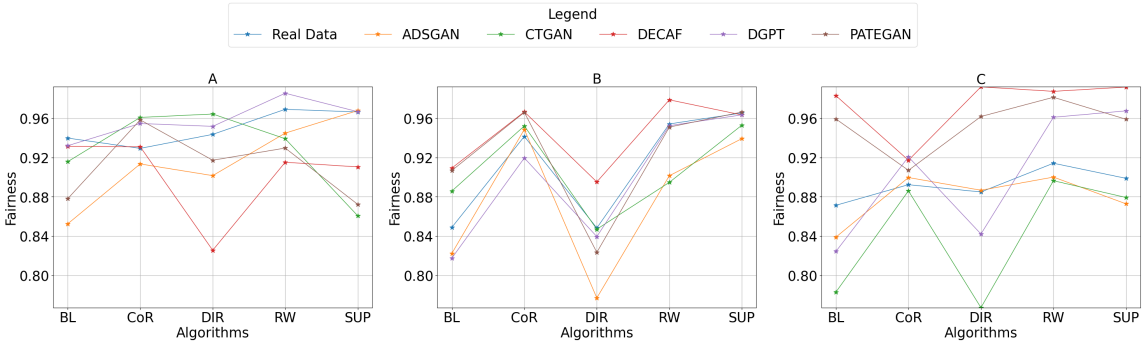

Fig. 2. The trend of fairness improvement or decline with respect to the baseline (BL) after the application of the 4 pre-processing fairness algorithms (i.e., CoR, DIR, RW, and SUP) for both real and synthetic datasets across all models. (Higher fairness is better) : $Fairness = \frac{3-|ABROCA|-|ERD|-|TPRD|}{3}$

Table 2. Percentage Change in Average Fairness Results Compared to Baseline Across Datasets and Synthetic Data for Dataset A, B, and C

Synthetic Data	Dataset A				Dataset B				Dataset C			
	CoR	DIR	RW	SUP	CoR	DIR	RW	SUP	CoR	DIR	RW	SUP
Real Data	-1.1%	0.4%	3.1%	2.8%	10.9%	-0.1%	12.4%	13.7%	2.4%	1.6%	4.9%	3.2%
ADSGAN	7.2%	**5.8%**	**10.8%**	**13.5%**	**15.4%**	-5.5%	9.6%	14.2%	7.2%	**5.7%**	7.3%	4.0%
CTGAN	4.9%	5.3%	2.6%	-6.0%	7.5%	-4.4%	1.0%	7.6%	**13.2%**	-2.0%	14.5%	12.3%
DECAF	-0.1%	-11.4%	-1.8%	-2.3%	6.3%	-1.6%	7.7%	6.0%	-6.7%	0.9%	0.5%	0.9%
DGPT	2.4%	2.1%	5.7%	3.7%	12.5%	**2.7%**	**16.5%**	**17.9%**	11.7%	2.2%	**16.6%**	**17.4%**
PATEGAN	**9.2%**	4.5%	5.9%	-0.7%	6.5%	-9.2%	4.9%	6.6%	-5.5%	0.3%	2.3%	-0.0%

Table 3. Average AUC Results Across Real and Synthetic Data for Dataset A with corresponding SEMs

Synthetic Data	BL	CoR	DIR	RW	SUP
Real Data	**0.94 ± 0.01**	**0.94 ± 0.02**	**0.97 ± 0.01**	**0.94 ± 0.04**	**0.95 ± 0.02**
ADSGAN	0.83 ± 0.01	0.92 ± 0.02	0.79 ± 0.01	0.74 ± 0.03	0.84 ± 0.02
CTGAN	0.76 ± 0.01	0.9 ± 0.01	0.86 ± 0.00	0.81 ± 0.02	0.77 ± 0.02
DECAF	0.51 ± 0.03	0.42 ± 0.06	0.49 ± 0.01	0.48 ± 0.02	0.58 ± 0.03
DGPT	0.92 ± 0.03	**0.94 ± 0.02**	0.94 ± 0.03	0.89 ± 0.07	0.91 ± 0.04
PATEGAN	0.58 ± 0.02	0.75 ± 0.02	0.57 ± 0.01	0.64 ± 0.02	0.60 ± 0.01

Table 4. Average AUC Results Across Real and Synthetic Data for Dataset B with corresponding SEMs

Synthetic Data	BL	CoR	DIR	RW	SUP
Real Data	0.63 ± 0.00	0.63 ± 0.01	**0.62 ± 0.00**	**0.64 ± 0.00**	0.61 ± 0.00
ADSGAN	**0.64 ± 0.00**	**0.63 ± 0.00**	0.61 ± 0.00	0.62 ± 0.00	**0.63 ± 0.00**
CTGAN	0.63 ± 0.01	0.63 ± 0.01	0.62 ± 0.00	0.61 ± 0.01	0.62 ± 0.00
DECAF	0.49 ± 0.00	0.56 ± 0.01	0.49 ± 0.00	0.47 ± 0.00	0.48 ± 0.01
DGPT	0.50 ± 0.01	0.50 ± 0.01	0.49 ± 0.01	0.50 ± 0.00	0.51 ± 0.00
PATEGAN	0.52 ± 0.00	0.51 ± 0.01	0.51 ± 0.01	0.54 ± 0.01	0.51 ± 0.01

Table 5. Average AUC Results Across Real and Synthetic Data for Dataset C with corresponding SEMs

Synthetic Data	BL	CoR	DIR	RW	SUP
Real Data	**0.84 ± 0.00**	**0.82 ± 0.00**	**0.84 ± 0.00**	**0.82 ± 0.00**	**0.83 ± 0.00**
ADSGAN	0.82 ± 0.00	0.81 ± 0.01	0.80 ± 0.00	0.79 ± 0.00	0.82 ± 0.00
CTGAN	0.74 ± 0.01	0.81 ± 0.01	0.71 ± 0.00	0.72 ± 0.01	0.74 ± 0.01
DECAF	0.65 ± 0.01	0.72 ± 0.01	0.67 ± 0.01	0.65 ± 0.01	0.65 ± 0.01
DGPT	0.65 ± 0.01	0.77 ± 0.04	0.63 ± 0.01	0.62 ± 0.01	0.62 ± 0.00
PATEGAN	0.57 ± 0.01	0.48 ± 0.08	0.55 ± 0.02	0.56 ± 0.01	0.56 ± 0.01

4 Individual Results for ML Models Before Pre-processing (i.e., Baseline Models) and After Pre-processing (i.e., Fair Models) Across Datasets

4.1 Same Train, Real Test

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.79	0.04	0.09	0.19
	LR	0.85	0.03	0.09	0.18
	RF	0.80	0.05	0.06	0.19
	NB	0.82	0.03	0.50	0.31
CTGAN	XGB	0.77	0.03	0.27	0.23
	LR	0.84	0.03	0.32	0.21
	RF	0.83	0.02	0.17	0.16
	NB	0.81	0.02	0.96	0.61
DECAF	XGB	0.59	0.04	0.00	0.21
	LR	0.76	0.05	0.01	0.20
	RF	0.75	0.05	0.00	0.22
	NB	0.75	0.05	0.05	0.22
PATEGAN	XGB	0.54	0.07	0.01	0.22
	LR	0.69	0.08	0.00	0.22
	RF	0.50	0.05	0.00	0.22
	NB	0.54	0.10	0.03	0.23
DGPT	XGB	0.65	0.09	0.02	0.22
	LR	0.83	0.02	0.00	0.22
	RF	0.81	0.03	0.00	0.22
	NB	0.81	0.01	0.82	0.51
Real	XGB	0.83	0.01	0.10	0.17
	LR	0.85	0.03	0.10	0.17
	RF	0.84	0.02	0.07	0.16
	NB	0.83	0.02	0.47	0.29

Table 6. Fairness and Performance results for all Models on Law School Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.81	0.04	0.07	0.18
	LR	0.85	0.01	0.06	0.16
	RF	0.82	0.04	0.06	0.18
	NB	0.83	0.02	0.23	0.18
CTGAN	XGB	0.82	0.03	0.15	0.15
	LR	0.85	0.01	0.13	0.15
	RF	0.84	0.02	0.14	0.15
	NB	0.83	0.01	0.29	0.20
DECAF	XGB	0.67	0.03	0.01	0.20
	LR	0.75	0.06	0.01	0.19
	RF	0.75	0.05	0.00	0.20
	NB	0.75	0.01	0.04	0.18
PATEGAN	XGB	0.54	0.07	0.00	0.20
	LR	0.72	0.07	0.00	0.20
	RF	0.29	0.02	0.00	0.20
	NB	0.38	0.17	0.00	0.20
DGPT	XGB	0.64	0.06	0.01	0.21
	LR	0.84	0.02	0.00	0.20
	RF	0.81	0.03	0.00	0.20
	NB	0.82	0.01	0.01	0.19
Real	XGB	0.83	0.04	0.09	0.18
	LR	0.85	0.02	0.09	0.15
	RF	0.85	0.02	0.07	0.16
	NB	0.83	0.01	0.25	0.19

Table 7. Fairness and Performance results for all Models on Law School Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.80	0.02	0.07	0.19
	LR	0.82	0.02	0.06	0.17
	RF	0.80	0.03	0.07	0.19
	NB	0.80	0.02	0.21	0.17
CTGAN	XGB	0.80	0.02	0.11	0.18
	LR	0.83	0.02	0.10	0.17
	RF	0.81	0.02	0.10	0.17
	NB	0.81	0.01	0.26	0.19
DECAF	XGB	0.69	0.02	0.01	0.21
	LR	0.74	0.06	0.01	0.20
	RF	0.73	0.03	0.00	0.21
	NB	0.73	0.02	0.03	0.20
PATEGAN	XGB	0.51	0.04	0.00	0.21
	LR	0.68	0.07	0.00	0.21
	RF	0.35	0.02	0.00	0.21
	NB	0.37	0.13	0.00	0.21
DGPT	XGB	0.66	0.04	0.00	0.21
	LR	0.82	0.02	0.00	0.21
	RF	0.80	0.02	0.00	0.21
	NB	0.80	0.03	0.00	0.20
Real	XGB	0.82	0.02	0.07	0.18
	LR	0.83	0.02	0.10	0.17
	RF	0.82	0.02	0.08	0.19
	NB	0.81	0.02	0.24	0.19

Table 8. Fairness and Performance results for all Models on Law School Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.77	0.02	0.00	0.14
	LR	0.82	0.03	0.00	0.15
	RF	0.81	0.01	0.00	0.16
	NB	0.81	0.03	0.21	0.16
CTGAN	XGB	0.75	0.01	0.07	0.14
	LR	0.82	0.03	0.01	0.14
	RF	0.82	0.02	0.05	0.12
	NB	0.82	0.03	0.28	0.20
DECAF	XGB	0.68	0.05	0.01	0.17
	LR	0.75	0.04	0.01	0.16
	RF	0.76	0.03	0.00	0.17
	NB	0.74	0.07	0.03	0.17
PATEGAN	XGB	0.59	0.13	0.00	0.17
	LR	0.74	0.05	0.00	0.17
	RF	0.40	0.04	0.00	0.17
	NB	0.38	0.23	0.00	0.17
DGPT	XGB	0.58	0.08	0.01	0.18
	LR	0.83	0.02	0.00	0.17
	RF	0.80	0.02	0.00	0.17
	NB	0.81	0.03	0.01	0.15
Real	XGB	0.80	0.04	0.02	0.14
	LR	0.82	0.03	0.01	0.15
	RF	0.83	0.04	0.01	0.15
	NB	0.82	0.03	0.24	0.18

Table 9. Fairness and Performance results for all Models on Law School Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.77	0.03	0.06	0.19
	LR	0.83	0.02	0.07	0.15
	RF	0.75	0.08	0.03	0.17
	NB	0.82	0.02	0.32	0.22
CTGAN	XGB	0.77	0.02	0.33	0.25
	LR	0.82	0.03	0.35	0.22
	RF	0.79	0.03	0.25	0.19
	NB	0.81	0.02	0.92	0.59
DECAF	XGB	0.69	0.04	0.04	0.20
	LR	0.76	0.05	0.02	0.17
	RF	0.77	0.03	0.00	0.19
	NB	0.75	0.05	0.06	0.19
PATEGAN	XGB	0.56	0.05	0.00	0.19
	LR	0.70	0.05	0.00	0.19
	RF	0.34	0.06	0.00	0.19
	NB	0.52	0.09	0.00	0.20
DGPT	XGB	0.72	0.13	0.05	0.22
	LR	0.82	0.03	0.00	0.19
	RF	0.79	0.01	0.00	0.19
	NB	0.81	0.02	0.81	0.51
Real	XGB	0.84	0.03	0.14	0.16
	LR	0.84	0.03	0.11	0.15
	RF	0.83	0.02	0.10	0.15
	NB	0.83	0.02	0.40	0.26

Table 10. Fairness and Performance results for all Models on Law School Dataset After DisparateImpactRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.53	0.05	0.14	0.12
	LR	0.49	0.06	0.13	0.12
	RF	0.51	0.08	0.05	0.02
	NB	0.54	0.02	0.99	0.17
CTGAN	XGB	0.62	0.05	0.20	0.09
	LR	0.62	0.03	0.14	0.06
	RF	0.62	0.04	0.21	0.11
	NB	0.62	0.02	0.94	0.16
DECAF	XGB	0.50	0.08	0.10	0.10
	LR	0.51	0.05	0.00	0.12
	RF	0.51	0.07	0.05	0.09
	NB	0.52	0.08	0.22	0.17
ADSGAN	XGB	0.62	0.11	0.08	0.15
	LR	0.63	0.02	0.23	0.06
	RF	0.63	0.05	0.04	0.12
	NB	0.63	0.05	0.90	0.16
DGPT	XGB	0.54	0.03	0.00	0.12
	LR	0.53	0.04	0.04	0.13
	RF	0.54	0.06	0.00	0.10
	NB	0.54	0.06	0.91	0.17
Real	XGB	0.61	0.04	0.15	0.10
	LR	0.63	0.02	0.16	0.09
	RF	0.62	0.03	0.16	0.10
	NB	0.63	0.03	0.70	0.11

Table 11. Fairness and Performance results for all Models on OULAD Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.50	0.07	0.00	0.06
	LR	0.42	0.07	0.00	0.06
	RF	0.48	0.03	0.01	0.06
	NB	0.43	0.04	0.03	0.03
CTGAN	XGB	0.62	0.03	0.05	0.05
	LR	0.64	0.03	0.03	0.05
	RF	0.62	0.03	0.03	0.04
	NB	0.64	0.02	0.07	0.05
DECAF	XGB	0.51	0.06	0.04	0.02
	LR	0.57	0.05	0.00	0.06
	RF	0.53	0.09	0.03	0.03
	NB	0.57	0.06	0.00	0.04
ADSGAN	XGB	0.60	0.03	0.03	0.02
	LR	0.63	0.04	0.06	0.06
	RF	0.62	0.03	0.04	0.02
	NB	0.63	0.05	0.09	0.05
DGPT	XGB	0.52	0.04	0.00	0.05
	LR	0.59	0.04	0.00	0.05
	RF	0.53	0.04	0.04	0.03
	NB	0.53	0.04	0.06	0.04
Real	XGB	0.64	0.05	0.02	0.03
	LR	0.64	0.04	0.05	0.05
	RF	0.63	0.05	0.06	0.03
	NB	0.62	0.04	0.08	0.05

Table 12. Fairness and Performance results for all Models on OULAD Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.49	0.08	0.04	0.01
	LR	0.51	0.05	0.00	0.04
	RF	0.50	0.05	0.00	0.02
	NB	0.52	0.05	0.04	0.02
CTGAN	XGB	0.59	0.06	0.02	0.06
	LR	0.64	0.05	0.03	0.01
	RF	0.64	0.05	0.00	0.04
	NB	0.65	0.05	0.17	0.08
DECAF	XGB	0.54	0.06	0.06	0.04
	LR	0.59	0.03	0.00	0.04
	RF	0.54	0.06	0.01	0.03
	NB	0.58	0.03	0.03	0.03
ADSGAN	XGB	0.62	0.03	0.05	0.02
	LR	0.64	0.03	0.02	0.02
	RF	0.64	0.04	0.12	0.05
	NB	0.64	0.05	0.17	0.07
DGPT	XGB	0.51	0.12	0.13	0.11
	LR	0.51	0.07	0.00	0.03
	RF	0.48	0.06	0.08	0.09
	NB	0.52	0.07	0.15	0.05
Real	XGB	0.61	0.06	0.13	0.04
	LR	0.65	0.04	0.03	0.00
	RF	0.64	0.04	0.07	0.02
	NB	0.64	0.05	0.15	0.06

Table 13. Fairness and Performance results for all Models on OULAD Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.49	0.09	0.02	0.01
	LR	0.46	0.06	0.00	0.03
	RF	0.48	0.07	0.03	0.02
	NB	0.48	0.05	0.00	0.03
CTGAN	XGB	0.59	0.04	0.01	0.02
	LR	0.60	0.08	0.05	0.01
	RF	0.59	0.07	0.02	0.00
	NB	0.60	0.06	0.05	0.00
DECAF	XGB	0.51	0.08	0.02	0.02
	LR	0.52	0.04	0.00	0.03
	RF	0.51	0.07	0.00	0.03
	NB	0.53	0.04	0.01	0.02
ADSGAN	XGB	0.57	0.03	0.00	0.01
	LR	0.61	0.06	0.03	0.01
	RF	0.61	0.06	0.02	0.02
	NB	0.60	0.05	0.09	0.01
DGPT	XGB	0.50	0.05	0.08	0.03
	LR	0.53	0.05	0.02	0.00
	RF	0.49	0.06	0.23	0.09
	NB	0.51	0.04	0.07	0.00
Real	XGB	0.60	0.11	0.00	0.02
	LR	0.60	0.08	0.02	0.01
	RF	0.59	0.11	0.09	0.01
	NB	0.59	0.05	0.08	0.01

Table 14. Fairness and Performance results for all Models on OULAD Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.52	0.05	0.04	0.03
	LR	0.47	0.04	0.19	0.12
	RF	0.50	0.04	0.14	0.12
	NB	0.52	0.04	0.99	0.21
CTGAN	XGB	0.58	0.03	0.17	0.06
	LR	0.62	0.07	0.07	0.01
	RF	0.60	0.05	0.09	0.05
	NB	0.62	0.05	0.89	0.19
DECAF	XGB	0.56	0.06	0.05	0.04
	LR	0.54	0.04	0.00	0.07
	RF	0.54	0.05	0.00	0.07
	NB	0.56	0.03	0.15	0.07
ADSGAN	XGB	0.55	0.08	0.10	0.09
	LR	0.61	0.07	0.14	0.05
	RF	0.59	0.05	0.10	0.09
	NB	0.61	0.05	0.81	0.17
DGPT	XGB	0.52	0.07	0.05	0.08
	LR	0.56	0.07	0.02	0.04
	RF	0.51	0.03	0.03	0.06
	NB	0.55	0.05	0.07	0.04
Real	XGB	0.60	0.05	0.17	0.03
	LR	0.62	0.06	0.10	0.01
	RF	0.62	0.05	0.14	0.02
	NB	0.61	0.05	0.84	0.17

Table 15. Fairness and Performance results for all Models on OULAD Dataset After DisparateImpactRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.96	0.04	0.02	0.03
	LR	0.98	0.07	0.02	0.02
	RF	0.97	0.04	0.02	0.03
	NB	0.89	0.09	0.09	0.14
CTGAN	XGB	0.91	0.08	0.04	0.07
	LR	0.95	0.04	0.09	0.12
	RF	0.95	0.06	0.05	0.07
	NB	0.88	0.11	0.03	0.15
DECAF	XGB	0.65	0.14	0.02	0.12
	LR	0.36	0.12	0.02	0.12
	RF	0.46	0.09	0.00	0.10
	NB	0.40	0.08	0.11	0.13
DGPT	XGB	0.97	0.05	0.02	0.08
	LR	0.98	0.10	0.04	0.02
	RF	0.95	0.08	0.01	0.02
	NB	0.86	0.08	0.10	0.10
PATEGAN	XGB	0.82	0.14	0.04	0.19
	LR	0.76	0.16	0.02	0.18
	RF	0.79	0.06	0.00	0.12
	NB	0.73	0.19	0.01	0.11
Real	XGB	0.99	0.08	0.02	0.05
	LR	0.99	0.03	0.07	0.04
	RF	0.97	0.02	0.02	0.04
	NB	0.91	0.12	0.12	0.20

Table 16. Fairness and Performance results for all Models on Maths Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.96	0.03	0.03	0.00
	LR	0.97	0.02	0.02	0.01
	RF	0.95	0.02	0.04	0.05
	NB	0.82	0.03	0.07	0.01
CTGAN	XGB	0.87	0.02	0.09	0.05
	LR	0.95	0.03	0.03	0.08
	RF	0.94	0.06	0.03	0.04
	NB	0.81	0.05	0.03	0.06
DECAF	XGB	0.45	0.06	0.05	0.07
	LR	0.40	0.20	0.00	0.06
	RF	0.50	0.20	0.02	0.05
	NB	0.40	0.23	0.07	0.16
DGPT	XGB	0.95	0.05	0.01	0.01
	LR	0.93	0.02	0.05	0.00
	RF	0.93	0.05	0.04	0.01
	NB	0.81	0.09	0.02	0.10
PATEGAN	XGB	0.79	0.14	0.04	0.17
	LR	0.68	0.05	0.03	0.07
	RF	0.74	0.16	0.00	0.06
	NB	0.62	0.04	0.06	0.06
Real	XGB	0.94	0.11	0.06	0.01
	LR	0.96	0.08	0.16	0.13
	RF	0.95	0.06	0.01	0.04
	NB	0.82	0.04	0.01	0.00

Table 17. Fairness and Performance results for all Models on Maths Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.92	0.07	0.10	0.04
	LR	0.96	0.05	0.07	0.06
	RF	0.93	0.06	0.08	0.01
	NB	0.88	0.18	0.17	0.17
CTGAN	XGB	0.88	0.06	0.06	0.03
	LR	0.92	0.04	0.01	0.05
	RF	0.92	0.04	0.00	0.07
	NB	0.87	0.13	0.08	0.08
DECAF	XGB	0.56	0.07	0.01	0.08
	LR	0.46	0.23	0.11	0.01
	RF	0.36	0.07	0.00	0.08
	NB	0.31	0.06	0.09	0.04
DGPT	XGB	0.95	0.04	0.08	0.06
	LR	0.97	0.04	0.10	0.07
	RF	0.95	0.07	0.08	0.06
	NB	0.88	0.03	0.07	0.02
PATEGAN	XGB	0.74	0.04	0.00	0.06
	LR	0.74	0.08	0.00	0.01
	RF	0.80	0.04	0.00	0.08
	NB	0.73	0.16	0.04	0.01
Real	XGB	0.97	0.04	0.09	0.07
	LR	0.97	0.04	0.09	0.04
	RF	0.96	0.07	0.14	0.08
	NB	0.88	0.12	0.06	0.02

Table 18. Fairness and Performance results for all Models on Maths Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.95	0.05	0.05	0.04
	LR	0.95	0.04	0.07	0.07
	RF	0.94	0.03	0.02	0.02
	NB	0.85	0.11	0.18	0.18
CTGAN	XGB	0.88	0.08	0.05	0.11
	LR	0.86	0.11	0.05	0.13
	RF	0.92	0.07	0.05	0.08
	NB	0.79	0.19	0.14	0.21
DECAF	XGB	0.55	0.04	0.03	0.14
	LR	0.33	0.19	0.00	0.18
	RF	0.47	0.07	0.06	0.21
	NB	0.29	0.28	0.01	0.23
DGPT	XGB	0.94	0.04	0.04	0.07
	LR	0.94	0.04	0.04	0.01
	RF	0.90	0.10	0.04	0.07
	NB	0.82	0.12	0.04	0.10
PATEGAN	XGB	0.62	0.07	0.07	0.11
	LR	0.69	0.12	0.01	0.20
	RF	0.62	0.17	0.02	0.21
	NB	0.67	0.15	0.03	0.23
Real	XGB	0.96	0.07	0.02	0.05
	LR	0.97	0.02	0.01	0.02
	RF	0.94	0.03	0.04	0.05
	NB	0.86	0.11	0.10	0.20

Table 19. Fairness and Performance results for all Models on Maths Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.89	0.10	0.09	0.02
	LR	0.94	0.06	0.03	0.02
	RF	0.93	0.05	0.03	0.04
	NB	0.86	0.07	0.02	0.07
CTGAN	XGB	0.94	0.04	0.00	0.03
	LR	0.95	0.04	0.00	0.04
	RF	0.91	0.03	0.00	0.05
	NB	0.85	0.03	0.01	0.02
DECAF	XGB	0.56	0.08	0.04	0.06
	LR	0.38	0.15	0.00	0.06
	RF	0.54	0.12	0.00	0.06
	NB	0.43	0.13	0.03	0.09
DGPT	XGB	0.94	0.04	0.06	0.05
	LR	0.94	0.01	0.17	0.09
	RF	0.93	0.14	0.06	0.05
	NB	0.76	0.15	0.08	0.16
PATEGAN	XGB	0.75	0.09	0.04	0.07
	LR	0.72	0.05	0.03	0.07
	RF	0.73	0.08	0.00	0.06
	NB	0.64	0.05	0.02	0.04
Real	XGB	0.96	0.04	0.03	0.04
	LR	0.97	0.03	0.08	0.10
	RF	0.95	0.06	0.06	0.04
	NB	0.89	0.06	0.01	0.04

Table 20. Fairness and Performance results for all Models on Maths Dataset After DisparateImpactRemover Algorithm was applied

4.2 Same Train, Same Test

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.83	0.08	0.07	0.14
	LR	0.82	0.06	0.09	0.15
	RF	0.82	0.08	0.07	0.15
	NB	0.81	0.08	0.57	0.39
CTGAN	XGB	0.75	0.02	0.26	0.15
	LR	0.74	0.03	0.34	0.15
	RF	0.75	0.03	0.19	0.15
	NB	0.73	0.04	0.94	0.30
DECAF	XGB	0.64	0.04	0.00	0.02
	LR	0.64	0.02	0.00	0.02
	RF	0.66	0.01	0.00	0.02
	NB	0.66	0.02	0.02	0.03
PATEGAN	XGB	0.53	0.09	0.00	0.04
	LR	0.57	0.05	0.00	0.04
	RF	0.59	0.10	0.00	0.04
	NB	0.59	0.10	0.01	0.04
DGPT	XGB	0.62	0.10	0.01	0.10
	LR	0.66	0.07	0.00	0.09
	RF	0.67	0.08	0.00	0.09
	NB	0.65	0.04	0.85	0.69
Real	XGB	0.84	0.03	0.07	0.15
	LR	0.84	0.01	0.07	0.14
	RF	0.84	0.02	0.06	0.15
	NB	0.83	0.01	0.50	0.32

Table 21. Fairness and Performance results for all Models on Law School Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.82	0.06	0.08	0.17
	LR	0.82	0.06	0.07	0.17
	RF	0.82	0.06	0.07	0.18
	NB	0.81	0.08	0.28	0.24
CTGAN	XGB	0.75	0.01	0.16	0.17
	LR	0.73	0.02	0.11	0.17
	RF	0.74	0.01	0.16	0.17
	NB	0.72	0.04	0.28	0.15
DECAF	XGB	0.64	0.02	0.00	0.01
	LR	0.65	0.01	0.00	0.01
	RF	0.67	0.01	0.00	0.01
	NB	0.66	0.01	0.00	0.01
PATEGAN	XGB	0.55	0.09	0.00	0.04
	LR	0.55	0.08	0.00	0.04
	RF	0.59	0.08	0.00	0.04
	NB	0.57	0.09	0.00	0.04
DGPT	XGB	0.63	0.04	0.00	0.07
	LR	0.61	0.03	0.00	0.07
	RF	0.63	0.03	0.00	0.07
	NB	0.62	0.03	0.00	0.07
Real	XGB	0.82	0.02	0.09	0.16
	LR	0.84	0.02	0.08	0.16
	RF	0.84	0.02	0.08	0.17
	NB	0.83	0.01	0.23	0.17

Table 22. Fairness and Performance results for all Models on Law School Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.80	0.02	0.07	0.19
	LR	0.82	0.02	0.06	0.17
	RF	0.80	0.03	0.07	0.19
	NB	0.80	0.02	0.21	0.17
CTGAN	XGB	0.80	0.02	0.11	0.18
	LR	0.83	0.02	0.10	0.17
	RF	0.81	0.02	0.10	0.17
	NB	0.81	0.01	0.26	0.19
DECAF	XGB	0.69	0.02	0.01	0.21
	LR	0.74	0.06	0.01	0.20
	RF	0.73	0.03	0.00	0.21
	NB	0.73	0.02	0.03	0.20
PATEGAN	XGB	0.51	0.04	0.00	0.21
	LR	0.68	0.07	0.00	0.21
	RF	0.35	0.02	0.00	0.21
	NB	0.37	0.13	0.00	0.21
DGPT	XGB	0.66	0.04	0.00	0.21
	LR	0.82	0.02	0.00	0.21
	RF	0.80	0.02	0.00	0.21
	NB	0.80	0.03	0.00	0.20
Real	XGB	0.82	0.02	0.07	0.18
	LR	0.83	0.02	0.10	0.17
	RF	0.82	0.02	0.08	0.19
	NB	0.81	0.02	0.24	0.19

Table 23. Fairness and Performance results for all Models on Law School Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.79	0.08	0.01	0.17
	LR	0.78	0.05	0.01	0.16
	RF	0.80	0.07	0.00	0.17
	NB	0.80	0.06	0.23	0.20
CTGAN	XGB	0.72	0.01	0.04	0.20
	LR	0.70	0.02	0.00	0.23
	RF	0.73	0.02	0.06	0.20
	NB	0.72	0.04	0.27	0.14
DECAF	XGB	0.63	0.02	0.00	0.01
	LR	0.65	0.03	0.00	0.01
	RF	0.67	0.03	0.00	0.01
	NB	0.67	0.04	0.00	0.01
PATEGAN	XGB	0.55	0.05	0.00	0.02
	LR	0.54	0.02	0.00	0.02
	RF	0.57	0.03	0.00	0.02
	NB	0.58	0.04	0.00	0.02
DGPT	XGB	0.60	0.06	0.00	0.07
	LR	0.61	0.04	0.00	0.07
	RF	0.64	0.04	0.00	0.07
	NB	0.62	0.04	0.00	0.07
Real	XGB	0.81	0.02	0.01	0.18
	LR	0.81	0.02	0.01	0.17
	RF	0.82	0.02	0.02	0.17
	NB	0.83	0.02	0.22	0.17

Table 24. Fairness and Performance results for all Models on Law School Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.79	0.02	0.06	0.15
	LR	0.81	0.03	0.09	0.15
	RF	0.79	0.01	0.03	0.16
	NB	0.80	0.04	0.35	0.26
CTGAN	XGB	0.70	0.05	0.30	0.16
	LR	0.72	0.03	0.37	0.16
	RF	0.71	0.04	0.26	0.16
	NB	0.71	0.03	0.93	0.31
DECAF	XGB	0.65	0.01	0.00	0.01
	LR	0.66	0.01	0.00	0.01
	RF	0.68	0.01	0.00	0.01
	NB	0.67	0.02	0.00	0.01
PATEGAN	XGB	0.48	0.02	0.01	0.06
	LR	0.56	0.05	0.00	0.05
	RF	0.57	0.08	0.00	0.05
	NB	0.58	0.10	0.00	0.05
DGPT	XGB	0.60	0.03	0.00	0.07
	LR	0.64	0.03	0.00	0.07
	RF	0.65	0.04	0.00	0.07
	NB	0.65	0.04	0.84	0.70
Real	XGB	0.84	0.01	0.12	0.14
	LR	0.84	0.01	0.09	0.14
	RF	0.83	0.01	0.08	0.15
	NB	0.83	0.02	0.36	0.24

Table 25. Fairness and Performance results for all Models on Law School Dataset After DisparateImpactRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.64	0.08	0.01	0.09
	LR	0.63	0.06	0.06	0.11
	RF	0.64	0.07	0.03	0.03
	NB	0.62	0.07	0.70	0.14
ADSGAN	XGB	0.65	0.04	0.24	0.07
	LR	0.65	0.04	0.27	0.08
	RF	0.64	0.02	0.22	0.04
	NB	0.63	0.06	0.90	0.16
DECAF	XGB	0.49	0.05	0.08	0.03
	LR	0.49	0.05	0.00	0.02
	RF	0.49	0.04	0.00	0.02
	NB	0.49	0.05	0.67	0.09
PATEGAN	XGB	0.51	0.05	0.00	0.10
	LR	0.51	0.06	0.00	0.08
	RF	0.53	0.07	0.00	0.10
	NB	0.52	0.06	0.51	0.09
DGPT	XGB	0.48	0.08	0.21	0.00
	LR	0.51	0.14	0.00	0.06
	RF	0.50	0.06	0.40	0.08
	NB	0.50	0.13	0.96	0.20
Real	XGB	0.62	0.03	0.18	0.07
	LR	0.64	0.03	0.11	0.05
	RF	0.63	0.04	0.20	0.05
	NB	0.62	0.04	0.84	0.17

Table 26. Fairness and Performance results for all Models on OULAD Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.62	0.03	0.04	0.03
	LR	0.62	0.03	0.07	0.04
	RF	0.62	0.03	0.06	0.04
	NB	0.61	0.04	0.10	0.07
ADSGAN	XGB	0.63	0.04	0.04	0.10
	LR	0.62	0.04	0.01	0.13
	RF	0.62	0.03	0.03	0.10
	NB	0.62	0.06	0.04	0.11
DECAF	XGB	0.50	0.10	0.04	0.02
	LR	0.47	0.08	0.00	0.04
	RF	0.49	0.06	0.00	0.03
	NB	0.47	0.07	0.01	0.04
PATEGAN	XGB	0.53	0.06	0.00	0.02
	LR	0.50	0.12	0.00	0.02
	RF	0.53	0.09	0.01	0.02
	NB	0.50	0.06	0.02	0.01
DGPT	XGB	0.50	0.09	0.02	0.02
	LR	0.51	0.06	0.02	0.02
	RF	0.51	0.04	0.02	0.02
	NB	0.51	0.08	0.02	0.02
Real	XGB	0.61	0.06	0.02	0.05
	LR	0.61	0.04	0.01	0.04
	RF	0.60	0.07	0.01	0.01
	NB	0.60	0.04	0.06	0.06

Table 27. Fairness and Performance results for all Models on OULAD Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
PATEGAN	XGB	0.49	0.08	0.04	0.01
	LR	0.51	0.05	0.00	0.04
	RF	0.50	0.05	0.00	0.02
	NB	0.52	0.05	0.04	0.02
CTGAN	XGB	0.59	0.06	0.02	0.06
	LR	0.64	0.05	0.03	0.01
	RF	0.64	0.05	0.00	0.04
	NB	0.65	0.05	0.17	0.08
DECAF	XGB	0.54	0.06	0.06	0.04
	LR	0.59	0.03	0.00	0.04
	RF	0.54	0.06	0.01	0.03
	NB	0.58	0.03	0.03	0.03
ADSGAN	XGB	0.62	0.03	0.05	0.02
	LR	0.64	0.03	0.02	0.02
	RF	0.64	0.04	0.12	0.05
	NB	0.64	0.05	0.17	0.07
DGPT	XGB	0.51	0.12	0.13	0.11
	LR	0.51	0.07	0.00	0.03
	RF	0.48	0.06	0.08	0.09
	NB	0.52	0.07	0.15	0.05
Real	XGB	0.61	0.06	0.13	0.04
	LR	0.65	0.04	0.03	0.00
	RF	0.64	0.04	0.07	0.02
	NB	0.64	0.05	0.15	0.06

Table 28. Fairness and Performance results for all Models on OULAD Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.62	0.12	0.11	0.09
	LR	0.60	0.10	0.07	0.11
	RF	0.62	0.15	0.12	0.09
	NB	0.60	0.08	0.12	0.10
ADSGAN	XGB	0.61	0.07	0.14	0.15
	LR	0.63	0.05	0.02	0.13
	RF	0.63	0.06	0.12	0.12
	NB	0.62	0.07	0.11	0.14
DECAF	XGB	0.49	0.05	0.00	0.01
	LR	0.47	0.03	0.00	0.02
	RF	0.47	0.04	0.00	0.02
	NB	0.47	0.05	0.04	0.00
PATEGAN	XGB	0.55	0.13	0.01	0.04
	LR	0.54	0.08	0.00	0.07
	RF	0.56	0.08	0.03	0.05
	NB	0.53	0.06	0.02	0.04
DGPT	XGB	0.49	0.10	0.05	0.07
	LR	0.49	0.06	0.00	0.04
	RF	0.50	0.04	0.00	0.04
	NB	0.50	0.13	0.01	0.04
Real	XGB	0.64	0.07	0.02	0.04
	LR	0.64	0.04	0.01	0.03
	RF	0.64	0.05	0.08	0.06
	NB	0.64	0.03	0.10	0.05

Table 29. Fairness and Performance results for all Models on OULAD Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.62	0.03	0.13	0.04
	LR	0.62	0.04	0.16	0.07
	RF	0.63	0.04	0.19	0.10
	NB	0.61	0.04	0.84	0.15
ADSGAN	XGB	0.61	0.05	0.34	0.11
	LR	0.62	0.06	0.35	0.11
	RF	0.60	0.07	0.34	0.12
	NB	0.62	0.06	0.89	0.18
DECAF	XGB	0.49	0.04	0.12	0.01
	LR	0.49	0.03	0.00	0.01
	RF	0.49	0.06	0.14	0.04
	NB	0.50	0.05	0.66	0.12
PATEGAN	XGB	0.53	0.03	0.23	0.07
	LR	0.50	0.13	0.03	0.03
	RF	0.52	0.04	0.26	0.05
	NB	0.50	0.16	0.97	0.12
DGPT	XGB	0.48	0.06	0.00	0.03
	LR	0.51	0.06	0.01	0.04
	RF	0.49	0.08	0.43	0.13
	NB	0.50	0.06	0.95	0.18
Real	XGB	0.62	0.09	0.12	0.03
	LR	0.62	0.06	0.18	0.03
	RF	0.62	0.06	0.13	0.01
	NB	0.61	0.06	0.89	0.22

Table 30. Fairness and Performance results for all Models on OULAD Dataset After DisparateImpactRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.76	0.06	0.09	0.07
	LR	0.75	0.18	0.00	0.08
	RF	0.76	0.11	0.02	0.10
	NB	0.78	0.20	0.15	0.17
DECAF	XGB	0.47	0.09	0.07	0.07
	LR	0.47	0.09	0.00	0.10
	RF	0.53	0.12	0.00	0.10
	NB	0.57	0.12	0.24	0.01
ADSGAN	XGB	0.82	0.09	0.14	0.09
	LR	0.84	0.16	0.24	0.16
	RF	0.86	0.09	0.18	0.11
	NB	0.80	0.15	0.19	0.18
PATEGAN	XGB	0.61	0.23	0.01	0.03
	LR	0.54	0.35	0.03	0.08
	RF	0.62	0.39	0.03	0.04
	NB	0.53	0.31	0.08	0.04
DGPT	XGB	0.97	0.03	0.10	0.01
	LR	0.94	0.06	0.16	0.06
	RF	0.92	0.07	0.13	0.03
	NB	0.84	0.09	0.10	0.01
Real	XGB	0.93	0.04	0.09	0.05
	LR	0.97	0.03	0.04	0.05
	RF	0.95	0.04	0.09	0.07
	NB	0.91	0.05	0.05	0.12

Table 31. Fairness and Performance results for all Models on Maths Dataset Before Fairness Algorithms were applied (i.e., Baseline Models)

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.77	0.08	0.17	0.08
	LR	0.80	0.16	0.12	0.14
	RF	0.80	0.16	0.10	0.12
	NB	0.72	0.16	0.20	0.18
DECAF	XGB	0.62	0.07	0.04	0.04
	LR	0.50	0.18	0.07	0.09
	RF	0.60	0.25	0.00	0.11
	NB	0.58	0.21	0.15	0.09
ADSGAN	XGB	0.87	0.04	0.03	0.03
	LR	0.86	0.10	0.04	0.09
	RF	0.84	0.04	0.03	0.01
	NB	0.79	0.04	0.07	0.06
PATEGAN	XGB	0.64	0.15	0.09	0.14
	LR	0.57	0.06	0.08	0.13
	RF	0.60	0.27	0.00	0.08
	NB	0.59	0.15	0.17	0.20
DGPT	XGB	0.94	0.03	0.02	0.03
	LR	0.97	0.02	0.01	0.03
	RF	0.95	0.03	0.07	0.00
	NB	0.78	0.07	0.03	0.16
Real	XGB	0.96	0.04	0.07	0.07
	LR	0.97	0.02	0.03	0.02
	RF	0.95	0.04	0.05	0.02
	NB	0.90	0.12	0.03	0.17

Table 32. Fairness and Performance results for all Models on Maths Dataset After Suppression Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
ADSGAN	XGB	0.92	0.07	0.10	0.04
	LR	0.96	0.05	0.07	0.06
	RF	0.93	0.06	0.08	0.01
	NB	0.88	0.18	0.17	0.17
CTGAN	XGB	0.88	0.06	0.06	0.03
	LR	0.92	0.04	0.01	0.05
	RF	0.92	0.04	0.00	0.07
	NB	0.87	0.13	0.08	0.08
DECAF	XGB	0.56	0.07	0.01	0.08
	LR	0.46	0.23	0.11	0.01
	RF	0.36	0.07	0.00	0.08
	NB	0.31	0.06	0.09	0.04
DGPT	XGB	0.95	0.04	0.08	0.06
	LR	0.97	0.04	0.10	0.07
	RF	0.95	0.07	0.08	0.06
	NB	0.88	0.03	0.07	0.02
PATEGAN	XGB	0.74	0.04	0.00	0.06
	LR	0.74	0.08	0.00	0.01
	RF	0.80	0.04	0.00	0.08
	NB	0.73	0.16	0.04	0.01
Real	XGB	0.97	0.04	0.09	0.07
	LR	0.97	0.04	0.09	0.04
	RF	0.96	0.07	0.14	0.08
	NB	0.88	0.12	0.06	0.02

Table 33. Fairness and Performance results for all Models on Maths Dataset After CorrelationRemover Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.78	0.12	0.04	0.03
	LR	0.85	0.07	0.10	0.02
	RF	0.83	0.06	0.07	0.06
	NB	0.79	0.05	0.17	0.01
DECAF	XGB	0.44	0.12	0.06	0.11
	LR	0.45	0.06	0.00	0.17
	RF	0.51	0.15	0.00	0.14
	NB	0.52	0.09	0.08	0.03
ADSGAN	XGB	0.76	0.05	0.12	0.10
	LR	0.77	0.05	0.06	0.07
	RF	0.76	0.06	0.03	0.01
	NB	0.66	0.05	0.12	0.05
PATEGAN	XGB	0.60	0.08	0.04	0.04
	LR	0.67	0.12	0.05	0.04
	RF	0.67	0.14	0.00	0.09
	NB	0.63	0.13	0.03	0.07
DGPT	XGB	0.97	0.03	0.01	0.01
	LR	0.96	0.03	0.10	0.04
	RF	0.94	0.04	0.05	0.03
	NB	0.69	0.05	0.04	0.03
Real	XGB	0.97	0.02	0.00	0.04
	LR	0.99	0.02	0.03	0.02
	RF	0.95	0.04	0.03	0.04
	NB	0.83	0.09	0.10	0.05

Table 34. Fairness and Performance results for all Models on Maths Dataset After Reweighing Algorithm was applied

	Model	AUC	ABROCA	TPRD	ERD
CTGAN	XGB	0.85	0.10	0.02	0.10
	LR	0.87	0.06	0.01	0.04
	RF	0.87	0.07	0.09	0.03
	NB	0.86	0.04	0.01	0.04
DECAF	XGB	0.52	0.11	0.00	0.06
	LR	0.46	0.26	0.25	0.25
	RF	0.50	0.18	0.08	0.06
	NB	0.49	0.33	0.29	0.21
ADSGAN	XGB	0.81	0.07	0.17	0.05
	LR	0.79	0.07	0.17	0.01
	RF	0.81	0.12	0.18	0.03
	NB	0.76	0.08	0.17	0.05
PATEGAN	XGB	0.59	0.21	0.01	0.01
	LR	0.56	0.24	0.00	0.03
	RF	0.58	0.22	0.00	0.03
	NB	0.55	0.21	0.07	0.07
DGPT	XGB	0.97	0.01	0.12	0.00
	LR	0.97	0.03	0.01	0.03
	RF	0.95	0.07	0.12	0.00
	NB	0.86	0.04	0.17	0.03
Real	XGB	0.98	0.02	0.09	0.06
	LR	0.98	0.02	0.11	0.09
	RF	0.98	0.02	0.09	0.04
	NB	0.95	0.02	0.06	0.05

Table 35. Fairness and Performance results for all Models on Maths Dataset After DisparateImpactRemover Algorithm was applied

5 Additional Figures

5.1 Additional privacy evaluation results

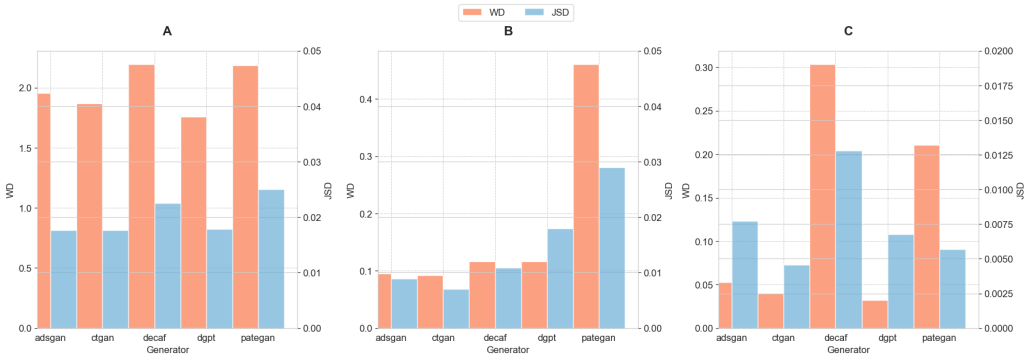

Fig. 3. distance and similarity metrics in privacy evaluation

Fig. 4. re-identification risk assessment metrics in privacy evaluation

5.2 Additional fairness evaluation results

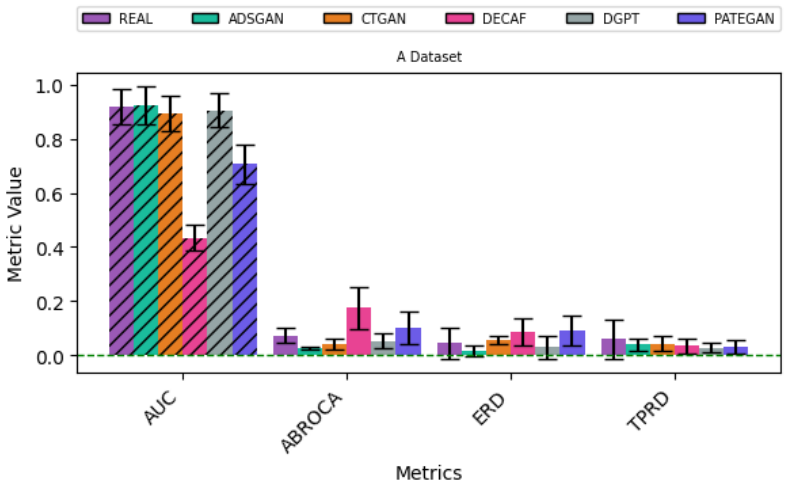

Fig. 5. Overall Fairness and AUC score for all Models for Dataset A. Bar heights denote average value of metric and the error bars indicate the standard deviation. Over here we just keep the fairness scores in their typical range without normalizing them (As reported in the main paper. Hence, fairness scores closer to 0 are the best and fairness scores closer to 1 are the worst. Note that we still Use the absolute values of the fairness scores. For the hatched AUC bars, the closer they are to 1 the better.

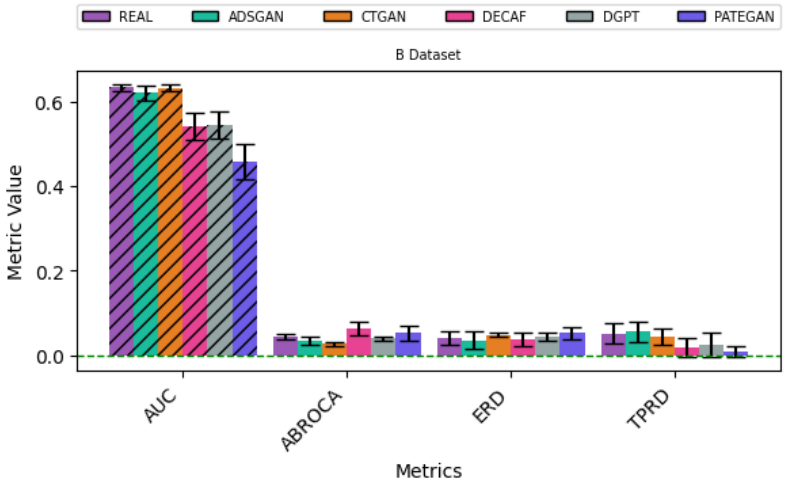

Fig. 6. Overall Fairness and AUC score for all Models for Dataset B. Bar heights denote average value of metric and the error bars indicate the standard deviation. Over here we just keep the fairness scores in their typical range without normalizing them (As reported in the main paper. Hence, fairness scores closer to 0 are the best and fairness scores closer to 1 are the worst. Note that we still Use the absolute values of the fairness scores. For the hatched AUC bars, the closer they are to 1 the better.

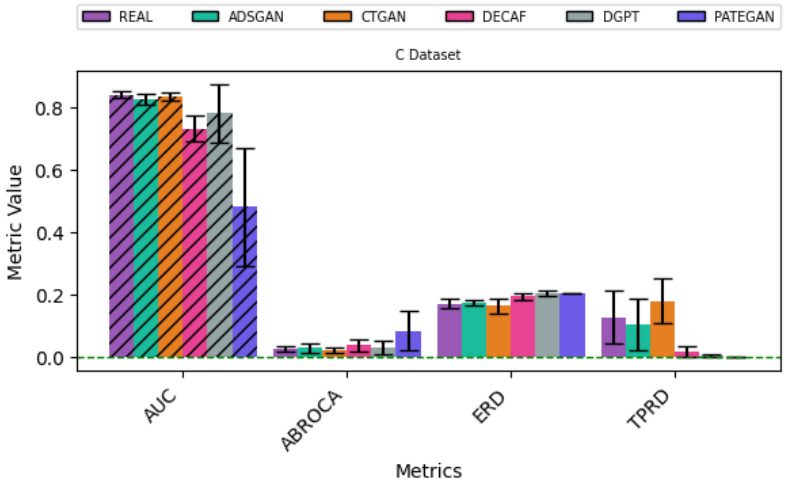

Fig. 7. Overall Fairness and AUC score for all Models for Dataset C. Bar heights denote average value of metric and the error bars indicate the standard deviation. Over here we just keep the fairness scores in their typical range without normalizing them (As reported in the main paper. Hence, fairness scores closer to 0 are the best and fairness scores closer to 1 are the worst. Note that we still Use the absolute values of the fairness scores. For the hatched AUC bars, the closer they are to 1 the better.

References

[1] Google. 2015. *Computing k-anonymity for a dataset.* [online] Available at: https://cloud.google.com/sensitive-data-protection/docs/compute-k-anonymity [Accessed: Month Day, Year].

[2] C. Horan. 2021. *Wasserstein Distance and Textual Similarity.* Neptune.ai, https://neptune.ai/blog/wasserstein-distance-and-textual-similarity [Accessed: Month Day, Year].

[3] F. Nielsen. 2019. On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy* 21, 5 (2019), 485. https://doi.org/10.3390/e21050485

[4] B. Van Breugel, H. Sun, Z. Qian, and M. Van Der Schaar. 2023. Membership Inference Attacks against Synthetic Data through Overfitting Detection. arXiv preprint, https://arxiv.org/pdf/2302.12580.