

Supplementary Information

Table S1: The default parameters in each deep generative model.

CTGAN		
Parameters	Explanations	Values
batch_size	Number of samples to process in each step when learning the algorithm	500
discriminator_dim	Size of the output samples for each one of the Discriminator Layers	(256, 256)
discriminator_decay	Weight decay for the Adam Optimizer in the discriminator	1e-6.
discriminator_lr	Learning rate (Discriminator)	2.00E-04
discriminator_steps	Number of discriminator updates to do for each generator update	1
embedding_dim	Size of the random sample passed to the Generator	128
generator_decay	Generator weight decay for the Adam Optimizer	1.00E-06
generator_dim	Size of the output samples for each one of the Residuals	(256, 256)
generator_lr	Learning rate (Generator)	2.00E-04
log_frequency	Whether to use log frequency of categorical levels in conditional sampling	TRUE
pac	Number of samples to group together when applying the discriminator	10
CopulaGAN		
batch_size	Number of data samples to process in each step when learning the algorithm	500
discriminator_dim	Size of the output samples for each one of the Discriminator Layers	(256, 256)
discriminator_decay	Discriminator weight decay for the Adam Optimizer	1e-6.
discriminator_lr	Learning rate (Discriminator)	2.00E-04
discriminator_steps	Number of discriminator updates to do for each generator update	1
embedding_dim	Size of the random sample passed to the Generator	128
generator_decay	Weight decay for the Adam Optimizer in the discriminator	1.00E-06
generator_dim	Size of the output samples for each one of the Residuals	(256, 256)
generator_lr	Learning rate (Generator)	2.00E-04
log_frequency	Whether to use log frequency of categorical levels in conditional sampling	TRUE
pac	Number of samples to group together when applying the discriminator	10
TVAE		
batch_size	Number of data samples to process in each step	500

compress_dims	Size of each hidden layer(Encoder)	(128, 128)
decompress_dims	Size of each hidden layer(Decoder)	(128, 128)
embedding_dim	Size of the random sample passed to the Generator	128
l2scale	L2 Regularization	1.00E-05
loss_factor	Multiplier for the reconstruction error	2
pac	Number of samples to group together when applying the discriminator	10

Table S2: Results of optimizing parameters.

number	value	params_colsample_bytree	params_gamma	params_learning_rate	params_max_depth	params_min_child_weight	params_n_estimators	params_reg_alpha	params_reg_lambda	params_subsample
0	0.800379	0.163051	8.89E-05	0.034841	1	4	147	5.89E-04	1.06E-03	0.222573
1	0.808356	0.028717	6.11E-06	0.079513	7	7	185	6.50E-04	2.07E-06	0.121946
2	1	0.275222	1.53E-08	0.015174	4	4	494	1.88E-03	7.37E-06	0.021813
3	0.800822	0.022496	4.57E-01	0.028486	6	10	457	9.00E-04	4.39E-03	0.236669
4	0.657322	0.030317	3.83E-05	0.028636	1	5	253	1.80E-06	7.28E-06	0.085671
5	0.808356	0.020041	1.41E-03	0.011766	7	9	429	2.12E-02	2.17E-07	0.166615
6	0.820148	0.018519	5.71E-07	0.052446	6	7	141	6.06E-06	3.33E-01	0.678972
7	0.980094	0.060789	1.09E-06	0.06629	5	3	244	1.53E-06	3.01E-05	0.018694
8	0.800379	0.18814	8.81E-07	0.030862	1	3	500	2.00E-06	2.52E-03	0.019702
9	0.812062	0.051189	5.79E-07	0.031905	4	2	446	1.86E-03	6.86E-08	0.034809
10	1	0.709559	2.61E-08	0.321826	9	1	359	9.06E-01	1.30E-08	0.038576
11	1	0.760085	1.23E-08	0.443292	9	1	365	5.02E-01	1.52E-08	0.045568
12	1	0.747126	2.82E-08	0.284557	3	1	336	2.47E-08	5.21E-07	0.013061
13	1	0.403796	1.22E-08	0.180683	9	6	353	8.06E-01	1.17E-08	0.048082
14	1	0.34135	2.75E-03	0.787026	3	4	53	5.22E-02	5.17E-05	0.027822
15	1	0.375008	1.92E-07	0.164227	8	2	404	3.01E-02	1.25E-01	0.01001
16	0.950131	0.191417	1.14E-07	0.011949	3	5	303	2.77E-05	2.41E-06	0.068489
17	1	0.304049	8.20E-03	0.873977	3	4	60	1.05E-02	1.46E-04	0.022799
18	0.959253	0.090854	8.65E-02	0.934162	4	7	51	5.72E-03	8.84E-04	0.020236
19	1	0.443988	1.20E-05	0.160144	8	3	406	4.74E-05	4.80E-01	0.011868
20	0.930723	0.513585	1.68E-03	0.143373	2	3	202	6.71E-05	2.56E-02	0.011249
21	1	0.993451	3.47E-02	0.488853	4	4	483	1.34E-01	4.52E-04	0.031156
22	1	0.54487	3.87E-04	0.359486	8	2	395	9.02E-08	1.51E-02	0.014927
23	0.808356	0.010232	6.50E-06	0.594751	8	1	307	2.40E-01	1.82E-04	0.057831
24	1	0.281485	7.12E-08	0.210272	9	6	338	7.46E-01	1.20E-08	0.054193
25	1	0.82146	3.98E-04	0.307003	5	1	402	1.84E-08	1.97E-02	0.014056
26	1	0.858941	3.70E-02	0.399896	5	2	465	7.53E-08	1.25E-02	0.016735
27	0.952818	0.136467	9.07E-08	0.209305	9	6	348	2.84E-07	1.40E-07	0.115757
28	1	0.288583	2.69E-04	0.10568	6	8	423	1.80E-04	6.43E-02	0.653358
29	1	0.863902	8.59E-03	0.639041	5	2	106	2.05E-08	6.85E-03	0.028713

Adopted parameters are highlighted in yellow, bold italic.

Table S3: Basic statistics

EEG signals			Empathetic dispositions		
EEG electrode	Mean	SD	Questionnaires	Mean	SD
AF3	0.23	9.13	Q1	3.91	1.70
F7	0.48	7.49	Q2	4.00	1.61
F3	0.23	8.00	Q3	5.00	0.89
FC5	0.41	6.94	Q4	4.45	0.69
T7	0.04	3.76	Q5	4.27	1.10
P7	-0.25	8.04	Q6	4.45	1.04
O1	0.41	6.84	Q7	4.91	0.83
O2	-0.05	6.55	Q8	4.82	0.87
P8	0.32	4.91	Q9	4.55	0.69
T8	0.07	5.98	Q10	5.00	0.77

FC6	0.25	8.04	Q11	3.45	1.51
F4	0.29	9.15	Q12	3.00	1.10
F8	0.16	9.09	Q13	3.09	0.94
AF4	0.48	11.91	Q14	3.73	1.10
---	---	---	Q15	3.64	0.81
---	---	---	Q16	3.64	1.12
---	---	---	Q17	3.00	0.89
---	---	---	Q18	3.36	1.29
---	---	---	Q19	3.00	1.10
---	---	---	Q20	2.82	1.54

Table S4: Data qualities in each deep generative model.

CTGAN										
MSE	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	7.5415	10.780 3	2.9204	8.5584	4.3942	3.2276	4.5944	1.2044
		7680 (200%)	7.3848	10.688 4	2.9178	8.7066	4.4468	3.2776	4.5034	1.1927
		11520 (300%)	7.4195	10.764 6	2.8891	8.7655	4.4580	3.2881	4.4833	1.1759
		19200 (500%)	7.4195	10.764 6	2.8891	8.7655	4.4580	3.2881	4.4833	1.1759
		38400 (1000%)	7.4195	10.764 6	2.8891	8.7655	4.4580	3.2881	4.4833	1.1759
		115200 (3000%)	7.4195	10.764 6	2.8891	8.7655	4.4580	3.2881	4.4833	1.1759
		Epoch		5	10	30	50	100	300	500
CC	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.4555	0.4433	0.6502	0.3737	0.5732	0.7894	0.6659	0.9600
		7680 (200%)	0.4647	0.4459	0.6510	0.3677	0.5715	0.7862	0.6679	0.9601
		11520 (300%)	0.4626	0.4444	0.6539	0.3649	0.5695	0.7871	0.6688	0.9610
		19200 (500%)	0.4626	0.4444	0.6539	0.3649	0.5695	0.7871	0.6688	0.9610

		<div>38400 (1000%)</div>	0.4626	0.4444	0.6539	0.3649	0.5695	0.7871	0.6688	0.9610
		<div>115200 (3000%)</div>	0.4626	0.4444	0.6539	0.3649	0.5695	0.7871	0.6688	0.9610
SMV	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	<div>3840 (100%)</div>	0.1001	0.0984	0.1000	0.0992	0.0812	0.0608	0.0426	0.0150
		<div>7680 (200%)</div>	0.0998	0.0988	0.1000	0.0992	0.0817	0.0610	0.0423	0.0148
		<div>11520 (300%)</div>	0.0999	0.0990	0.1002	0.0993	0.0818	0.0612	0.0422	0.0151
		<div>19200 (500%)</div>	0.0999	0.0990	0.1002	0.0993	0.0818	0.0612	0.0422	0.0151
		<div>38400 (1000%)</div>	0.0999	0.0990	0.1002	0.0993	0.0818	0.0612	0.0422	0.0151
		<div>115200 (3000%)</div>	0.0999	0.0990	0.1002	0.0993	0.0818	0.0612	0.0422	0.0151
CopulaGAN										
MSE	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	<div>3840 (100%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
		<div>7680 (200%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
		<div>11520 (300%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
		<div>19200 (500%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
		<div>38400 (1000%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
		<div>115200 (3000%)</div>	7.656	6.062	2.543	3.052	4.730	3.323	3.762	0.600
CC	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	<div>3840 (100%)</div>	0.6520	0.6951	0.8399	0.6915	0.5884	0.6979	0.8308	0.9326
		<div>7680</div>	0.6520	0.6951	0.8399	0.6915	0.5884	0.6979	0.8308	0.9326

		(3000%)								
CC	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.8883	0.8940	0.9294	0.9631	0.9658	0.9864	0.9880	0.9850
		7680 (200%)	0.8884	0.8933	0.9277	0.9635	0.9653	0.9907	0.9892	0.9870
		11520 (300%)	0.8883	0.8934	0.9276	0.9636	0.9648	0.9911	0.9897	0.9868
		19200 (500%)	0.8883	0.8934	0.9276	0.9636	0.9648	0.9911	0.9897	0.9868
		38400 (1000%)	0.8883	0.8934	0.9276	0.9636	0.9648	0.9911	0.9897	0.9868
		115200 (3000%)	0.8883	0.8934	0.9276	0.9636	0.9648	0.9911	0.9897	0.9868
	Epoch		5	10	30	50	100	300	500	1000
SMV	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	---	---	---	0.0735	0.0394	0.0051	0.0052	0.0037
		7680 (200%)	---	---	---	0.0733	0.0392	0.0053	0.0050	0.0036
		11520 (300%)	---	---	---	0.0730	0.0393	0.0052	0.0048	0.0035
		19200 (500%)	---	---	---	0.0730	0.0393	0.0052	0.0048	0.0035
		38400 (1000%)	---	---	---	0.0730	0.0393	0.0052	0.0048	0.0035
		115200 (3000%)	---	---	---	0.0730	0.0393	0.0052	0.0048	0.0035
	Epoch		5	10	30	50	100	300	500	1000

Abbreviations: MSE; mean squared error, CC; correlation coefficient, SMV; structure matching value

Table S5: Predictive performance in each deep generative model.

CTGAN										
Accuracy	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation)	3840 (100%)	0.501	0.478	0.477	0.62	0.892	0.82	0.961	1

	ratio)	7680 (200%)	0.579	0.555	0.568	0.582	0.949	0.868	0.96	1
		11520 (300%)	0.513	0.53	0.379	0.54	0.895	0.805	0.948	1
		19200 (500%)	0.513	0.53	0.379	0.54	0.895	0.805	0.948	1
		38400 (1000%)	0.513	0.53	0.379	0.54	0.895	0.805	0.948	1
		115200 (3000%)	0.513	0.53	0.379	0.54	0.895	0.805	0.948	1
AUC	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.484	0.484	0.467	0.651	0.971	0.875	0.997	1
		7680 (200%)	0.608	0.575	0.582	0.639	1	0.917	0.995	1
		11520 (300%)	0.514	0.54	0.346	0.554	0.992	0.836	0.991	1
		19200 (500%)	0.514	0.54	0.346	0.554	0.992	0.836	0.991	1
		38400 (1000%)	0.514	0.54	0.346	0.554	0.992	0.836	0.991	1
		115200 (3000%)	0.514	0.54	0.346	0.554	0.992	0.836	0.991	1
Precision	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.501	0.476	0.479	0.602	1	0.882	0.94	1
		7680 (200%)	0.588	0.548	0.556	0.562	1	0.973	0.956	1
		11520 (300%)	0.517	0.536	0.388	0.543	1	0.947	0.92	1
		19200 (500%)	0.517	0.536	0.388	0.543	1	0.947	0.92	1
		38400 (1000%)	0.517	0.536	0.388	0.543	1	0.947	0.92	1
		115200	0.517	0.536	0.388	0.543	1	0.947	0.92	1

		(3000%)								
Recall	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.591	0.432	0.531	0.706	0.784	0.74	0.984	1
		7680 (200%)	0.529	0.622	0.677	0.745	0.898	0.758	0.964	1
		11520 (300%)	0.398	0.445	0.419	0.513	0.789	0.646	0.982	1
		19200 (500%)	0.398	0.445	0.419	0.513	0.789	0.646	0.982	1
		38400 (1000%)	0.398	0.445	0.419	0.513	0.789	0.646	0.982	1
		115200 (3000%)	0.398	0.445	0.419	0.513	0.789	0.646	0.982	1
F1	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.542	0.453	0.504	0.65	0.879	0.805	0.962	1
		7680 (200%)	0.557	0.583	0.61	0.641	0.947	0.852	0.96	1
		11520 (300%)	0.45	0.486	0.403	0.527	0.882	0.768	0.95	1
		19200 (500%)	0.45	0.486	0.403	0.527	0.882	0.768	0.95	1
		38400 (1000%)	0.45	0.486	0.403	0.527	0.882	0.768	0.95	1
		115200 (3000%)	0.45	0.486	0.403	0.527	0.882	0.768	0.95	1
CopulaGAN										
Accuracy	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999
		7680 (200%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999
		11520 (300%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999

		19200 (500%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999
		38400 (1000%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999
		115200 (3000%)	0.526	0.487	0.4	0.422	0.762	0.88	0.974	0.999
AUC	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
		7680 (200%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
		11520 (300%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
		19200 (500%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
		38400 (1000%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
		115200 (3000%)	0.524	0.465	0.369	0.385	0.833	0.947	0.998	1
Precision	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
		7680 (200%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
		11520 (300%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
		19200 (500%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
		38400 (1000%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
		115200 (3000%)	0.52	0.489	0.42	0.414	0.802	0.865	0.992	1
Recall	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes	3840 (100%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997

	(Augmentation ratio)	7680 (200%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997
		11520 (300%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997
		19200 (500%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997
		38400 (1000%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997
		115200 (3000%)	0.661	0.568	0.523	0.378	0.695	0.901	0.956	0.997
F1	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
		7680 (200%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
		11520 (300%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
		19200 (500%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
		38400 (1000%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
		115200 (3000%)	0.583	0.525	0.466	0.395	0.745	0.883	0.973	0.999
TVAE										
Accuracy	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.505	0.538	0.622	0.665	0.993	0.997	0.984	1
		7680 (200%)	0.566	0.526	0.618	0.617	0.965	0.987	0.986	1
		11520 (300%)	0.457	0.487	0.646	0.604	0.961	0.987	0.993	0.997
		19200 (500%)	0.457	0.487	0.646	0.604	0.961	0.987	0.993	0.997
		38400 (1000%)	0.457	0.487	0.646	0.604	0.961	0.987	0.993	0.997

		115200 (3000%)	0.457	0.487	0.646	0.604	0.961	0.987	0.993	0.997
AUC	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.507	0.548	0.658	0.746	1	1	0.999	1
		7680 (200%)	0.579	0.511	0.651	0.667	0.999	1	0.999	1
		11520 (300%)	0.449	0.48	0.69	0.65	0.994	0.999	1	1
		19200 (500%)	0.449	0.48	0.69	0.65	0.994	0.999	1	1
		38400 (1000%)	0.449	0.48	0.69	0.65	0.994	0.999	1	1
		115200 (3000%)	0.449	0.48	0.69	0.65	0.994	0.999	1	1
Precision	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.506	0.533	0.626	0.674	0.995	0.997	0.982	1
		7680 (200%)	0.557	0.522	0.625	0.624	0.936	0.977	0.975	1
		11520 (300%)	0.468	0.488	0.63	0.601	0.973	0.987	0.992	1
		19200 (500%)	0.468	0.488	0.63	0.601	0.973	0.987	0.992	1
		38400 (1000%)	0.468	0.488	0.63	0.601	0.973	0.987	0.992	1
		115200 (3000%)	0.468	0.488	0.63	0.601	0.973	0.987	0.992	1
Recall	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.466	0.602	0.609	0.641	0.992	0.997	0.987	1
		7680 (200%)	0.654	0.615	0.591	0.589	0.997	0.977	0.997	1
		11520 (300%)	0.633	0.536	0.706	0.622	0.948	0.987	0.995	0.995

		19200 (500%)	0.633	0.536	0.706	0.622	0.948	0.987	0.995	0.995
		38400 (1000%)	0.633	0.536	0.706	0.622	0.948	0.987	0.995	0.995
		115200 (3000%)	0.633	0.536	0.706	0.622	0.948	0.987	0.995	0.995
F1	Epoch		5	10	30	50	100	300	500	1000
	Synthesized sample sizes (Augmentation ratio)	3840 (100%)	0.485	0.565	0.617	0.657	0.993	0.997	0.984	1
		7680 (200%)	0.601	0.565	0.608	0.606	0.966	0.987	0.986	1
		11520 (300%)	0.538	0.511	0.666	0.611	0.96	0.987	0.993	0.997
		19200 (500%)	0.538	0.511	0.666	0.611	0.96	0.987	0.993	0.997
		38400 (1000%)	0.538	0.511	0.666	0.611	0.96	0.987	0.993	0.997
		115200 (3000%)	0.538	0.511	0.666	0.611	0.96	0.987	0.993	0.997

Table S6: Computation time.

Epoch	5	10	30	50	100	300	500	1000
CTGAN	13.9s	15.3s	31.1s	47s	85s	225s	354s	869s
CopulaGAN	14.6s	16.3s	30.3s	38.8s	85s	219s	417s	860s
TVAE	13.6s	12.4s	18.6s	23.2s	39.1s	85s	132s	258s

Code S1(A): CTGAN

Importing related modules

```
import numpy as np
```

```
import pandas as pd
```

```
from ctgan import CTGAN
```

#Prepare CTGAN algorithm

```
ctgan = CTGAN(epochs=30)
```

#Learn CTGAN algorithm

```
ctgan.fit(data, discrete_columns)
```

#Generate synthesized data by learned algorithm

#num_ samples is the number of synthesized data

```
synthetic_data = ctgan.sample(n_samples)
```

Code S1(B): CopulaGAN

Importing related modules

```
import numpy as np
```

```
import pandas as pd
```

```
from sdv.single_table import CopulaGANSynthesizer
```

```
from sdv.metadata.single_table import SingleTableMetadata
```

#Setting metadata

```
metadata = SingleTableMetadata()
```

```
metadata.detect_from_dataframe(data)
```

#Prepare CopulaGAN algorithm

```
synthesizer = CopulaGANSynthesizer(metadata, epochs=n)
```

#Learn CopulaGAN algorithm

```
synthesizer.fit(data)
```

#Generate synthesized data by learned algorithm

#num_row is the number of synthesized data

```
synthetic_data = synthesizer.sample(num_rows=num_rows)
```

Code S1(C): TVAE

Importing related modules

```
import numpy as np
```

```
import pandas as pd
```

```
from sdv.tabular import TVAE
```

#Prepare TVAE algorithm

```
tvae = TVAE(epochs=n)
```

#Learn TVAE algorithm

```
tvae.fit(data)
```

#Generate synthesized data by learned algorithm

#num_row is the number of synthesized data

```
synthetic_data = tvae.sample(num_rows)
```

Code S1(D): XGBoost(Hyperparameter tuning)

Importing related modules

```
import optuna
```

```
import xgboost as xgb
```

Build XGBoost instance

```
X_data, y_data = preprocessing(data, category_dict)
```

```
xgtrain = xgb.DMatrix(X_data, y_data)
```

```
def objective(trial):
```

Candidates for hyperparameters tuning

```
    params = {
```

```
        'max_depth': trial.suggest_int('max_depth', 1, 9),
```

```
        'learning_rate': trial.suggest_loguniform('learning_rate', 0.01, 1.0),
```

```
        'n_estimators': trial.suggest_int('n_estimators', 50, 500),
```

```
        'min_child_weight': trial.suggest_int('min_child_weight', 1, 10),
```

```
        'gamma': trial.suggest_loguniform('gamma', 1e-8, 1.0),
```

```
        'subsample': trial.suggest_loguniform('subsample', 0.01, 1.0),
```

```
        'colsample_bytree': trial.suggest_loguniform('colsample_bytree', 0.01, 1.0),
```

```
        'reg_alpha': trial.suggest_loguniform('reg_alpha', 1e-8, 1.0),
```

```
        'reg_lambda': trial.suggest_loguniform('reg_lambda', 1e-8, 1.0),
```

```
        'eval_metric': 'auc'
```

```
    }
```

Cross-validation

```
    cv_results = xgb.cv(
```

```
        params,
```

```
        xgtrain,
```

```
        num_boost_round=100,
```

```
        seed=0,
```

```

        nfold=5, # The number of cross-validation
        metrics={"auc"},
        early_stopping_rounds=5
    )

    return cv_results["test-auc-mean"].min()

```

```

study = optuna.create_study(direction="maximize")
study.optimize(objective, n_trials=30, timeout=300)

```

Code S2(A): XGBoost(Validation:Setting the trained model)

Importing related modules

```

from xgboost import XGBClassifier
from sklearn.model_selection import GridSearchCV
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
from sklearn.metrics import accuracy_score
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
from sklearn.metrics import f1_score
from sklearn.metrics import confusion_matrix

```

Prediction

```

def learn_predict(X, y, X_test, y_test):
    xgb=XGBClassifier(learning_rate=0.243258,max_depth=6,min_child_weight=1,gamma=0,colsample_bytree=0.76,n_estimators=365,reg_alpha=0.50,reg_lambda=0,subsample=0.045)

```

```

    xgb.fit(X, y)

```

```

    predictions = xgb.predict_proba(X_test)

```

Calculate and print the evaluation metrics

```

    auc = roc_auc_score(y_test, predictions[:, 1])
    bool_prediction = (predictions[:, 1] >= 0.5).astype(int)
    acc = accuracy_score(y_test, bool_prediction)
    precision = precision_score(y_test, bool_prediction)

```

```

recall = recall_score(y_test, bool_pediction)
f1 = f1_score(y_test, bool_pediction)
print("AUC: {:.3f}".format(auc))
print("Accuracy {:.3f}".format(acc))
print("Precision: {:.3f}".format(precision))
print("Recall: {:.3f}".format(recall))
print("f1: {:.3f}".format(f1))
print("Confusion matrix:")
print(confusion_matrix(y_test, bool_pediction))
return (auc, acc, precision, recall, f1)

```

Code S2(B): XGBoost(Validation:Predicted results)

```

n_samples = [3840,7680,11520,19200,38400,115200] # Generated data
auc_list0 = []
acc_list0 = []
precision_list0 = []
recall_list0 = []
f1_list0 = []

for n in n_samples:
    print("==" * 12)
    print(" # of samples: ", n)
    print("==" * 12)
    ac = learn_predict(X0_syn[:n], y0_syn[:n], X_test, y_test)
    print()
    auc_list0.append(ac[0])
    acc_list0.append(ac[1])
    precision_list0.append(ac[2])
    recall_list0.append(ac[3])
    f1_list0.append(ac[4])
    print("--" * 12)
    print(" # of samples: ", n)
    print("--" * 12)
    st = np.mean(synthetic_data[:n],axis=0)
    print(st)

```



```
print("//" * 12)
print(" # of samples: ", n)
print("//" * 12)
res=synthetic_data[:n].corr()
pd.set_option('display.max_columns', 100)
print(res)
```