Appendix F

Full Results of Perturbation Effectiveness Testing

F.1 Using Tabular Data

Table F.1: The effectiveness of perturbations in identifying feature importance to the model when using tabular data, using the MAPE metric to measure the difference between the original model output and the model output for the perturbed set of data points. Brighter colour indicates a higher correlation between the true model importance and the perturbation-based rankings.

Dataset	Model	Baseline Zero	Baseline Max	Baseline Mean	Baseline Min	Feature Permutation
Breast Cancer	Decision Tree	0.4912	0.2606	0.2824	0.4912	0.6885
	Logit	0.5010	0.5416	0.5299	0.5020	0.6454
	Naïve Bayes	0.0580	0.2675	0.4468	0.0659	0.3259
COMPAS	Decision Tree	0.5998	0.7608	0.6318	0.5998	0.9159
	Logit	0.1913	0.7092	0.4025	0.1884	0.9463
	Naïve Bayes	0.1210	0.0794	0.4505	0.1208	0.2933
Diabetes	Decision Tree	0.3879	0.6078	0.4201	0.3879	0.8037
	Logit	0.5025	0.7321	0.5493	0.5018	0.8743
	Naïve Bayes	0.1220	0.0579	0.4421	0.1213	0.2443
Adult Income	Decision Tree	0.5689	0.7857	0.5857	0.5689	0.9718
	Logit	0.0487	0.7493	0.2030	0.0487	0.8408
	Naïve Bayes	0.2725	0.3293	0.7620	0.2725	0.4841
Iris	Decision Tree	0.5000	0.5000	0.5000	0.5000	1.0000
	Logit	0.2622	0.3333	0.7179	0.1410	0.7949
	Naïve Bayes	0.1084	0.2692	0.9359	0.0128	0.8590
Mushroom	Decision Tree	0.1475	0.3895	0.1475	0.1475	0.5032
	Logit	0.0205	0.6095	0.3809	0.0205	0.8675
	Naïve Bayes	0.0234	0.0553	0.0487	0.0234	0.0637
Nursery	Decision Tree	0.5000	0.5000	0.5000	0.5000	1.0000
	Logit	0.0942	0.5207	0.8154	0.0942	0.9862
	Naïve Bayes	0.1950	0.4316	0.8287	0.1950	0.8977

Table F.2: The effectiveness of perturbations in identifying feature importance to the model when using tabular data, using the MAPE metric to measure the difference between the original model output and the model output for the perturbed set of data points. Brighter colour indicates a higher correlation between the true model importance and the perturbation-based rankings.

Dataset	Model	Baseline Zero	Baseline Max	Baseline Mean	Baseline Min	Feature Permutation
Breast Cancer	Decision Tree	0.4912	0.2606	0.2824	0.4912	0.6893
	Logit	0.5010	0.5416	0.5299	0.5020	0.6564
	Naïve Bayes	0.0580	0.2673	0.4468	0.0659	0.3298
COMPAS	Decision Tree	0.5998	0.7608	0.6318	0.5998	0.9136
	Logit	0.1913	0.7092	0.4025	0.1884	0.9488
	Naïve Bayes	0.1210	0.0798	0.4505	0.1208	0.2874
Diabetes	Decision Tree	0.3879	0.6078	0.4201	0.3879	0.8133
	Logit	0.5025	0.7321	0.5493	0.5018	0.8764
	Naïve Bayes	0.1220	0.0579	0.4421	0.1213	0.2879
Adult Income	Decision Tree	0.5689	0.7857	0.5857	0.5689	0.9717
	Logit	0.0487	0.7493	0.2030	0.0487	0.8416
	Naïve Bayes	0.2725	0.3293	0.7620	0.2725	0.4792
Iris	Decision Tree	0.5000	0.5000	0.5000	0.5000	1.0000
	Logit	0.2622	0.3333	0.7179	0.1410	0.7949
	Naïve Bayes	0.1084	0.2692	0.9359	0.0128	0.8590
Mushroom	Decision Tree	0.1475	0.3895	0.1475	0.1475	0.5032
	Logit	0.0205	0.6095	0.3809	0.0205	0.8629
	Naïve Bayes	0.0234	0.0553	0.0487	0.0234	0.0634
Nursery	Decision Tree	0.5000	0.5000	0.5000	0.5000	1.0000
	Logit	0.0942	0.5207	0.8154	0.0942	0.9744
	Naïve Bayes	0.1950	0.4316	0.8287	0.1950	0.8954

F.2 Using Event Logs

Table F.3: The effectiveness of perturbations in identifying feature importance to the model when using event log data, using the RMSE metric to measure the difference between the original model output and the model output for the perturbed set of data points. Brighter colour indicates a higher correlation between the true model importance and the perturbation-based rankings.

Preprocessing	Dataset	Model	Baseline Zero	Baseline Max	Baseline Mean	Baseline Min	Feature Permutation
Single & Aggregate	BPIC2012	Decision Tree	0.5982	0.4018	0.4018	0.5982	0.9196
		Logit	-0.2037	0.9057	0.0454	-0.2037	0.9343
		Naïve Bayes	0.1494	0.5457	0.4814	0.1494	0.7324
	Production	Decision Tree	0.5841	0.6901	0.5620	0.5841	0.9856
		Logit	0.0545	0.5699	0.4166	0.0545	0.5613
		Naïve Bayes	-0.0413	0.3281	0.3245	-0.0413	0.3341
	Sepsis Cases	Decision Tree	0.8304	0.1696	0.1696	0.8304	1.0000
		Logit	-0.1036	0.7098	0.2352	-0.1036	0.8072
		Naïve Bayes	-0.0563	0.3538	0.0610	-0.0563	0.3632
Prefix & Aggregate	BPIC2012	Decision Tree	0.5640	0.7614	0.3514	0.5640	0.9635
		Logit	0.0871	0.7785	0.2757	0.0872	0.8941
		Naïve Bayes	0.1645	0.5027	0.3695	0.1645	0.6314
	Production	Decision Tree	0.4389	0.6875	0.4382	0.4389	0.9724
		Logit	0.1583	0.6381	0.5616	0.1583	0.6910
		Naïve Bayes	0.0123	0.3254	0.1740	0.0123	0.3539
	Sepsis Cases	Decision Tree	0.3903	0.7823	0.2894	0.3903	0.9924
		Logit	0.0939	0.5923	0.4158	0.0947	0.7068
		Naïve Bayes	0.0265	0.4117	0.2757	0.0313	0.4454
Prefix & Index	BPIC2012	Decision Tree	0.5314	0.6683	0.5511	0.5314	0.9376
		Logit	0.1660	0.7899	0.4399	0.1660	0.9131
		Naïve Bayes	0.0903	0.5792	0.4481	0.0903	0.6443
	Production	Decision Tree	0.3922	0.7250	0.3165	0.3922	0.9725
		Logit	0.0679	0.6917	0.5451	0.0679	0.7771
		Naïve Bayes	0.0086	0.1296	0.0567	0.0086	0.1277
	Sepsis Cases	Decision Tree	0.5276	0.6939	0.4588	0.5276	0.9576
		Logit	0.0429	0.5552	0.3839	0.0451	0.6937
		Naïve Bayes	0.0331	0.3458	0.2012	0.0334	0.3815

Table F.4: The effectiveness of perturbations in identifying feature importance to the model when using event log data, using the MAPE metric to measure the difference between the original model output and the model output for the perturbed set of data points. Brighter colour indicates a higher correlation between the true model importance and the perturbation-based rankings.

Preprocessing	Dataset	Model	Baseline Zero	Baseline Max	Baseline Mean	Baseline Min	Feature Permutation
Single & Aggregate	BPIC2012	Decision Tree	0.5982	0.4018	0.4018	0.5982	0.9196
		Logit	-0.2037	0.9057	0.0454	-0.2037	0.9445
		Naïve Bayes	0.1494	0.5457	0.4814	0.1494	0.7268
	Production	Decision Tree	0.5841	0.6901	0.5620	0.5841	0.9859
		Logit	0.0545	0.5699	0.4166	0.0545	0.5544
		Naïve Bayes	-0.0413	0.3281	0.3245	-0.0413	0.3332
	Sepsis Cases	Decision Tree	0.8304	0.1696	0.1696	0.8304	1.0000
		Logit	-0.1036	0.7098	0.2352	-0.1036	0.8049
		Naïve Bayes	-0.0563	0.3538	0.0610	-0.0563	0.3593
Prefix & Aggregate	BPIC2012	Decision Tree	0.5640	0.7614	0.3514	0.5640	0.9640
		Logit	0.0871	0.7785	0.2757	0.0872	0.8987
		Naïve Bayes	0.1645	0.5027	0.3695	0.1645	0.6399
	Production	Decision Tree	0.4389	0.6874	0.4382	0.4389	0.9726
		Logit	0.1582	0.6381	0.5616	0.1582	0.6853
		Naïve Bayes	0.0123	0.3254	0.1740	0.0123	0.3561
	Sepsis Cases	Decision Tree	0.3903	0.7823	0.2894	0.3903	0.9924
		Logit	0.0939	0.5924	0.4158	0.0947	0.6998
		Naïve Bayes	0.0265	0.4117	0.2757	0.0313	0.4447
Prefix & Index	BPIC2012	Decision Tree	0.5314	0.6683	0.5511	0.5314	0.9367
		Logit	0.1660	0.7899	0.4399	0.1660	0.9239
		Naïve Bayes	0.0903	0.5792	0.4481	0.0903	0.6497
	Production	Decision Tree	0.3922	0.7250	0.3165	0.3922	0.9724
		Logit	0.0679	0.6917	0.5451	0.0679	0.7893
		Naïve Bayes	0.0086	0.1296	0.0567	0.0086	0.1280
	Sepsis Cases	Decision Tree	0.5276	0.6939	0.4588	0.5276	0.9576
		Logit	0.0429	0.5552	0.3839	0.0451	0.6834
		Naïve Bayes	0.0331	0.3458	0.2012	0.0334	0.3809