Spplementary Materials for the Article: gbt-HIPS: Explaining the Classifications of Gradient Boosted Tree Ensembles

Julian Hatwell

February 2021

Table 1: Mean Coverage of Explanations of the Stochastic GBT Model

data	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	0.40 ± 0.01	0.18 ± 0.00	0.08 ± 0.01	0.21 ± 0.00	$\mathbf{0.73} \pm 0.01$	0.47 ± 0.01
bank	$\mathbf{0.75} \pm 0.01$	0.32 ± 0.00	0.44 ± 0.02	0.14 ± 0.00	0.70 ± 0.01	0.11 ± 0.00
car	0.28 ± 0.01	0.20 ± 0.01	0.29 ± 0.01	0.32 ± 0.01	$\mathbf{0.49} \pm 0.01$	0.21 ± 0.01
cardio	0.44 ± 0.01	0.06 ± 0.00	$\mathbf{0.72} \pm 0.02$	0.33 ± 0.00	0.47 ± 0.01	0.50 ± 0.01
credit	0.39 ± 0.01	0.35 ± 0.01	$\mathbf{0.60} \pm 0.02$	0.49 ± 0.01	0.48 ± 0.01	0.46 ± 0.01
german	0.28 ± 0.01	0.09 ± 0.01	$\mathbf{0.66} \pm 0.02$	0.35 ± 0.00	0.31 ± 0.01	0.23 ± 0.01
lending	0.51 ± 0.01	0.11 ± 0.00	N/A	$\mathbf{0.88} \pm 0.01$	N/A	0.47 ± 0.01
nursery	0.16 ± 0.00	0.13 ± 0.00	0.27 ± 0.01	0.20 ± 0.00	$\mathbf{0.56} \pm 0.01$	0.15 ± 0.00
rcdv	$\mathbf{0.48} \pm 0.01$	0.19 ± 0.00	0.35 ± 0.02	0.20 ± 0.00	0.22 ± 0.01	0.19 ± 0.01

Table 2: Mean Exclusive Coverage of Explanations of the Stochastic GBT Model

data	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	**0.36** ± 0.01	0.17 ± 0.00	0.03 ± 0.00	0.20 ± 0.00	0.30 ± 0.00	0.26 ± 0.01
bank	**0.48** ± 0.01	0.29 ± 0.00	0.00 ± 0.00	0.13 ± 0.00	0.37 ± 0.00	0.10 ± 0.00
car	**0.26** ± 0.00	0.20 ± 0.01	0.19 ± 0.01	0.24 ± 0.00	0.20 ± 0.01	0.21 ± 0.01
cardio	**0.40** ± 0.01	0.06 ± 0.00	0.01 ± 0.00	0.26 ± 0.01	0.39 ± 0.01	0.29 ± 0.01
credit	0.38 ± 0.01	0.35 ± 0.01	0.22 ± 0.01	**0.42** ± 0.00	0.32 ± 0.01	0.41 ± 0.01
german	0.21 ± 0.01	0.09 ± 0.01	0.11 ± 0.01	**0.24** ± 0.01	0.19 ± 0.00	0.16 ± 0.01
lending	**0.50** ± 0.01	0.11 ± 0.00	N/A	0.06 ± 0.00	N/A	0.35 ± 0.01
nursery	0.16 ± 0.00	0.13 ± 0.00	0.13 ± 0.00	**0.19** ± 0.00	0.22 ± 0.00	0.14 ± 0.00
rcdv	**0.31** ± 0.01	0.18 ± 0.00	0.00 ± 0.00	0.17 ± 0.00	0.00 ± 0.00	0.18 ± 0.00

Table 3: Mean Ranks of Exclusive Coverage of Explanations of the Stochastic GBT Model

The overall position of gbt-HIPS is given in the *pos* column.

data	N	M	pos	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	1000	6	1st	**1.95**	4.24	5.48	3.69	2.34	3.31
bank	1000	6	1st	**1.54**	2.76	5.96	4.08	2.24	4.43
car	517	6	2nd	3.17	3.62	4.03	**3.00**	3.58	3.61
cardio	637	6	1st	**1.92**	4.91	5.72	3.38	2.21	2.86
credit	206	6	3rd	3.25	3.32	5.36	2.59	4.28	**2.20**
german	299	6	2nd	2.80	4.52	4.72	**2.31**	3.17	3.48
lending	631	4	1st	**1.23**	3.22	N/A	3.54	N/A	2.01
nursery	1000	6	3rd	3.28	4.43	3.44	3.17	**2.70**	3.98
rcdv	1000	6	1st	**1.50**	2.90	5.53	2.82	5.39	2.87

Table 4: Mean Precision of Explanations of the Stochastic GBT Model

data	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	0.95 ± 0.00	**0.96** ± 0.00	0.58 ± 0.01	0.85 ± 0.01	0.73 ± 0.01	0.79 ± 0.01
bank	**0.95** ± 0.00	**0.95** ± 0.00	0.39 ± 0.01	0.86 ± 0.01	0.87 ± 0.01	0.69 ± 0.01
car	0.95 ± 0.00	**0.97** ± 0.01	0.85 ± 0.01	0.57 ± 0.02	0.76 ± 0.01	0.93 ± 0.01
cardio	0.94 ± 0.00	0.95 ± 0.01	0.61 ± 0.01	0.73 ± 0.02	0.82 ± 0.01	0.80 ± 0.01
credit	0.98 ± 0.00	0.99 ± 0.00	0.65 ± 0.02	0.81 ± 0.02	0.69 ± 0.01	0.89 ± 0.01
german	**0.85** ± 0.01	0.82 ± 0.02	0.58 ± 0.02	0.52 ± 0.02	0.60 ± 0.01	0.74 ± 0.02
lending	**0.98** ± 0.00	**0.98** ± 0.01	N/A	0.65 ± 0.01	N/A	0.90 ± 0.01
nursery	0.92 ± 0.00	**0.97** ± 0.01	0.85 ± 0.01	0.71 ± 0.01	0.33 ± 0.01	0.90 ± 0.01
rcdv	0.89 ± 0.00	**0.95** ± 0.00	0.20 ± 0.01	0.67 ± 0.00	0.04 ± 0.00	0.84 ± 0.01

Table 5: Mean Reliability of Explanations of the Stochastic GBT Model

data	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	**0.95** ± 0.00	**0.94** ± 0.00	0.64 ± 0.01	0.85 ± 0.01	0.72 ± 0.01	0.83 ± 0.01
bank	**0.95** ± 0.00	**0.95** ± 0.00	0.56 ± 0.01	0.86 ± 0.01	0.87 ± 0.01	0.78 ± 0.01
car	**0.93** ± 0.00	0.89 ± 0.01	0.83 ± 0.01	0.56 ± 0.02	0.75 ± 0.01	0.88 ± 0.01
cardio	**0.92** ± 0.00	0.86 ± 0.01	0.60 ± 0.01	0.72 ± 0.01	0.81 ± 0.01	0.78 ± 0.01
credit	**0.96** ± 0.00	0.95 ± 0.01	0.64 ± 0.02	0.78 ± 0.02	0.68 ± 0.01	0.87 ± 0.01
german	**0.80** ± 0.01	0.74 ± 0.01	0.57 ± 0.02	0.52 ± 0.02	0.58 ± 0.02	0.71 ± 0.01
lending	**0.96** ± 0.00	0.93 ± 0.00	N/A	0.64 ± 0.01	N/A	0.89 ± 0.01
nursery	**0.90** ± 0.00	0.88 ± 0.01	0.83 ± 0.01	0.70 ± 0.01	0.33 ± 0.01	0.84 ± 0.01
rcdv	0.88 ± 0.00	**0.93** ± 0.00	0.38 ± 0.01	0.67 ± 0.01	0.29 ± 0.00	0.87 ± 0.01

Table 6: Mean Ranks of Reliability of Explanations of the Stochastic GBT Model

The overall position of gbt-HIPS is given in the *pos* column.

data	N	M	pos	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	1000	6	3rd	2.68	**2.19**	4.99	2.62	5.12	3.39
bank	1000	6	4th	3.20	**2.23**	5.85	2.89	3.99	2.83
car	517	6	2nd	2.61	**2.60**	2.97	5.52	4.50	2.81
cardio	637	6	1st	**1.95**	3.54	5.60	2.58	3.46	3.87
credit	206	6	2nd	1.94	**1.69**	5.36	4.37	4.57	3.06
german	299	6	1st	**2.18**	2.56	4.40	4.64	4.22	3.00
lending	631	4	1st	**1.23**	2.18	N/A	3.97	N/A	2.62
nursery	1000	6	3rd	2.73	2.61	**2.28**	4.66	5.55	3.17
rcdv	1000	6	3rd	2.78	**1.74**	5.19	3.97	5.44	1.89

Table 7: 0.75 Reliability Floor for Explanations of the SAMME AdaBoost Model

The overall position of gbt-HIPS is given in the *pos* column.

data	pos	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	1st	**1.00**	0.95	0.45	0.92	0.83	0.84
bank	1st	**0.97**	0.96	0.44	0.83	0.90	0.69
car	1st	**1.00**	0.78	0.76	0.47	0.66	0.81
cardio	1st	**0.94**	0.86	0.69	0.65	0.77	0.76
credit	1st	**0.99**	0.97	0.45	0.87	0.43	0.95
german	1st	**0.67**	0.62	0.29	0.37	0.43	0.54
lending	1st	**0.99**	0.94	N/A	0.73	N/A	0.88
nursery	1st	**0.91**	0.84	0.75	0.66	0.00	0.76
rcdv	1st	**0.99**	0.95	0.00	0.57	0.00	0.84

Table 8: Mean Rule Length of Explanations of the Stochastic GBT Model

data	gbt-HIPS	Anchors	BRL	defragTrees	inTrees	LORE
adult	1.33 ± 0.03	2.12 ± 0.06	13.10 ± 0.21	4.73 ± 0.03	0.34 ± 0.04	3.20 ± 0.05
bank	1.71 ± 0.03	1.49 ± 0.07	5.74 ± 0.17	3.39 ± 0.02	1.12 ± 0.03	4.90 ± 0.06
car	1.73 ± 0.05	2.10 ± 0.06	2.86 ± 0.14	2.00 ± 0.00	3.76 ± 0.09	1.98 ± 0.06
cardio	1.51 ± 0.04	3.00 ± 0.08	1.83 ± 0.10	7.47 ± 0.13	1.06 ± 0.05	2.16 ± 0.05
credit	1.54 ± 0.04	1.72 ± 0.06	0.38 ± 0.05	1.95 ± 0.08	1.24 ± 0.03	1.23 ± 0.05
german	1.68 ± 0.08	3.07 ± 0.12	1.23 ± 0.05	3.14 ± 0.07	2.53 ± 0.08	2.13 ± 0.08
lending	1.19 ± 0.02	2.03 ± 0.04	N/A	6.80 ± 0.03	N/A	1.89 ± 0.04
nursery	2.24 ± 0.04	2.90 ± 0.06	7.21 ± 0.26	2.00 ± 0.00	2.38 ± 0.05	2.48 ± 0.05
rcdv	1.59 ± 0.03	2.60 ± 0.05	5.12 ± 0.12	3.16 ± 0.03	1.51 ± 0.02	4.51 ± 0.06