Appendix A. Experiment results for all tested ensemble sizes

Table A.1. Results presented in BCE for homogeneous decision tree ensembles using BNG_credit-a dataset.

Programming language	Ensemble size / Weighting	2	3	5	8	13
Python	Equal	0.323	0.329	0.322	0.319	0.319
	Mono	0.338	0.339	0.339	0.339	0.339
	Perf	0.323	0.328	0.322	0.319	0.318
	PosShap	0.322	0.328	0.320	0.318	0.317
	Rand	0.329	0.334	0.322	0.32	0.319
R	Equal	0.408	0.407	0.385	0.378	0.368
	Mono	0.408	0.408	0.408	0.408	0.408
	Perf	0.408	0.407	0.384	0.376	0.367
	PosShap	0.408	0.407	0.379	0.370	0.360
	Rand	0.408	0.407	0.389	0.38	0.371

Table A.2. Results presented in BCE for homogeneous logistic regression ensembles using Bank Marketing dataset

Programming language	Ensemble size / Weighting	2	3	5	8	13
Python	Equal	0.239	0.241	0.240	0.239	0.243
	maxShap	0.239	0.240	0.239	0.27	0.416
	Mono	0.239	0.240	0.239	0.239	0.240
	Perf	0.239	0.241	0.239	0.238	0.237
	posShap	0.239	0.240	0.239	0.238	0.238
	Rand	0.240	0.241	0.240	0.240	0.245
R	Equal	0.241	0.240	0.240	0.240	0.246
	maxShap	0.241	0.239	0.240	0.288	0.441
	Mono	0.241	0.239	0.240	0.24	0.24
	Perf	0.241	0.240	0.240	0.239	0.236
	posShap	0.241	0.239	0.240	0.239	0.238
	Rand	0.241	0.240	0.241	0.241	0.247

Table A.1. Results presented in BCE for homogeneous logistic regression ensembles using BNG_credit-a dataset

Programming language	Ensemble size / Weighting	2	3	5	8	13
Python	Equal	0.327	0.326	0.326	0.326	0.326
	Mono	0.326	0.326	0.326	0.326	0.326
	Perf	0.327	0.326	0.326	0.326	0.326
	posShap	0.327	0.326	0.326	0.326	0.326
	Rand	0.327	0.326	0.327	0.326	0.327
R	Equal	0.327	0.327	0.327	0.327	0.327
	Mono	0.327	0.327	0.326	0.327	0.327
	Perf	0.327	0.327	0.327	0.327	0.327
	posShap	0.327	0.327	0.327	0.327	0.327
	Rand	0.327	0.327	0.327	0.327	0.327

Table A.2. Results presented in BCE for homogeneous decision tree ensembles using Bank Marketing dataset

Programming language	Ensemble size / Weighting	2	3	5	8	13
Python	Equal	0.260	0.260	0.256	0.252	0.289
	maxShap	0.261	0.275	0.325	0.344	0.366
	Mono	0.272	0.273	0.272	0.27	0.282
	Perf	0.260	0.259	0.254	0.251	0.287
	posShap	0.260	0.261	0.257	0.254	0.274
	Rand	0.263	0.262	0.258	0.256	0.289
R	Equal	0.280	0.278	0.269	0.265	0.266
	maxShap	0.280	0.278	0.272	0.281	0.291
	Mono	0.282	0.283	0.282	0.282	0.283
	Perf	0.280	0.278	0.269	0.266	0.267
	posShap	0.280	0.278	0.270	0.266	0.264
	Rand	0.280	0.279	0.271	0.267	0.268

Table A.3. Results presented in BCE for heterogeneous ensembles trained using BNG_credit-a dataset

Programming language	Ensemble size / Weighting	4	6	10	16
Python	Equal	0.313	0.318	0.314	0.313
	maxShap	0.313	0.318	0.313	0.312
	Mono	0.326	0.326	0.326	0.326
	Perf	0.313	0.318	0.314	0.313
	posShap	0.313	0.318	0.313	0.312
	Rand	0.316	0.318	0.314	0.313
R	Equal	0.348	0.348	0.342	0.339
	maxShap	0.348	0.343	0.341	0.341
	Mono	0.327	0.326	0.326	0.326
	Perf	0.348	0.347	0.342	0.340
	posShap	0.348	0.343	0.341	0.341
	Rand	0.348	0.346	0.342	0.339

Table A.4. Results presented in BCE of heterogeneous ensembles using Bank Marketing dataset

Programming language	Ensemble size / Weighting	4	6	10	16
Python	Equal	0.236	0.237	0.235	0.233
	maxShap	0.239	0.244	0.26	0.269
	Mono	0.239	0.24	0.239	0.239
	Perf	0.237	0.239	0.238	0.237
	posShap	0.239	0.24	0.238	0.235
	Rand	0.237	0.238	0.236	0.235
R	Equal	0.236	0.237	0.235	0.233
	maxShap	0.239	0.244	0.26	0.269
	Mono	0.241	0.239	0.239	0.24
	Perf	0.237	0.239	0.238	0.237
	posShap	0.239	0.24	0.238	0.235
	Rand	0.237	0.238	0.236	0.235