Beta calibration: a well-founded and easily implemented improvement on logistic calibration for binary classifiers – Supplementary material

Meelis Kull
University of Bristol
University of Tartu

Telmo de Menezes e Silva Filho
Universidade Federal de Pernambuco
Centro de Informática

Peter Flach
University of Bristol

This material supplements the AISTATS 2017 paper on beta calibration and presents tables and critical difference diagrams for all results obtained in the experimental analysis. In all the tables, best results are marked in **bold** and subscript numbers indicate the ranks. Due to space limitations, numbers are rounded to three decimal digits, therefore, differences that occur after the third digit are not shown.

We evaluated the effect of applying beta calibration and its variations to the scores produced by Naive Bayes and Adaboost on 41 datasets from UCI, see Table 1 with the detailed information. Multiclass datasets were transformed into binary by calling the biggest class positive and all the other classes negative. We compared the performance of beta calibration, beta[a=b] calibration, beta[m=1/2] calibration, isotonic calibration, logistic calibration and uncalibrated probabilities, in terms of Brier score (BS) and log-loss (LL) and accuracy.

The results were obtained from 10 times 5-fold cross-validation, totalling 50 executions. Within each execution we used 3-fold internal cross-validation with 2 folds for learning the model and 1 for fitting the calibration map. Thus, three calibrated classifiers were generated during each execution, the outputs of these three were averaged to provide predictions on the test fold. The same methodology was used in the paper proposing the logistic calibration (Platt, 2000). All experiments were written in Python and the code is publicly available[1].

For Naive Bayes (NB), we used the implementation provided by Scikit-learn (Pedregosa et al., 2011). For boosting we used 200 decision stumps as weak learners and implemented two different versions of the standard Adaboost algorithm. The first is the original Adaboost with probabilities extracted in the standard way as in (Friedman et al., 2000), we refer to it as Ada-O. The second is the one implemented in Scikit-learn's based on Adaboost method SAMME (Zhu et al., 2009), we refer to it as Ada-S.

We first compared the full 3-parameter beta calibration with logistic and isotonic calibration methods, as well as with the uncalibrated probabilities, across all 3x3 settings (NB, Ada-O, Ada-S; LL, BS, Accuracy). The results for NB, Ada-O and Ada-S are respectively provided in Tables 2-4, 8-10, 14-16, and the respective critical difference diagrams in Figures 1-3, 7-9, 13-15. We then continued to compare the variants of beta calibration, leaving the non-parametric isotonic out of the picture. The results are provided in Tables 5-7, 11-13, 17-19, and the respective critical difference diagrams in Figures 4-6, 10-12, 16-18.

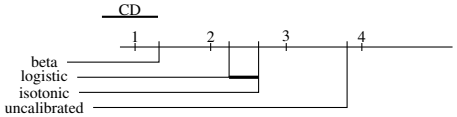

Figure 1: Critical difference diagram for log-loss results with Naive Bayes as base classifier. Friedman test shows significance at p-value=$6.879e-17$.

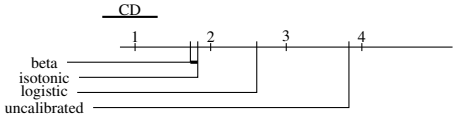

Figure 2: Critical difference diagram for Brier score results with Naive Bayes as base classifier. Friedman test shows significance at p-value=$1.002e-14$.

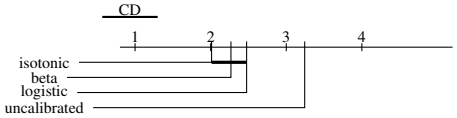

Figure 3: Critical difference diagram for accuracy results with Naive Bayes as base classifier. Friedman test shows significance at p-value=$3.201e-05$.

[1] https://betacal.github.io

Table 1: Description of the 41 classification datasets from UCI used for the experiments.

Name	Samples	Features	Classes
abalone	4177	8	3
autos	159	25	6
balance-scale	625	4	3
car	1728	6	4
cleveland	297	13	5
credit-approval	653	15	2
dermatology	358	34	6
diabetes	768	8	2
ecoli	336	7	8
flare	1389	10	6
german	1000	20	2
glass	214	9	6
heart-statlog	270	13	2
hepatitis	155	19	2
horse	300	27	2
ionosphere	351	34	2
iris	150	4	3
landsat-satellite	6435	36	6
letter	35000	16	26
libras-movement	360	90	15
lung-cancer	96	7129	2
mfeat-karhunen	2000	64	10
mfeat-morphological	2000	6	10
mfeat-zernike	2000	47	10
mushroom	8124	22	2
optdigits	5620	64	10
page-blocks	5473	10	5
pendigits	10992	16	10
scene-classification	2407	294	2
segment	2310	19	7
shuttle	101500	9	7
sonar	208	60	2
spambase	4601	57	2
tic-tac	958	9	2
vehicle	846	18	4
vowel	990	10	11
waveform-5000	5000	40	3
wdbc	569	30	2
wpbc	194	33	2
yeast	1484	8	10
zoo	101	16	7

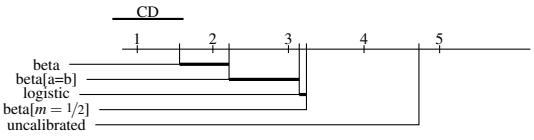

Figure 4: Critical difference diagram for log-loss results of parametric methods with Naive Bayes as base classifier. Friedman test shows significance at p-value=$1.202e{-}20$.

Table 2: Log-loss results for Naive Bayes (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $6.879e{-}17$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	1.170_4	$\mathbf{0.621}_1$	0.626_3	0.623_2
	(0.077)	(**0.008**)	(0.017)	(0.008)
autos	0.810_4	$\mathbf{0.529}_1$	0.548_2	0.558_3
	(0.305)	(**0.043**)	(0.172)	(0.037)
balance	0.284_4	$\mathbf{0.117}_1$	0.157_3	0.118_2
	(0.013)	(**0.033**)	(0.121)	(0.032)
car	0.396_3	0.382_2	$\mathbf{0.382}_1$	0.396_4
	(0.015)	(0.022)	(**0.044**)	(0.026)
clevela	0.631_4	$\mathbf{0.415}_1$	0.605_3	0.424_2
	(0.220)	(**0.074**)	(0.385)	(0.076)
credit-	1.043_4	$\mathbf{0.390}_1$	0.539_3	0.448_2
	(0.332)	(**0.051**)	(0.237)	(0.055)
dermato	0.282_4	$\mathbf{0.060}_1$	0.081_3	0.062_2
	(0.340)	(**0.047**)	(0.118)	(0.047)
diabete	0.612_4	$\mathbf{0.502}_1$	0.570_3	0.516_2
	(0.116)	(**0.032**)	(0.163)	(0.034)
ecoli	1.376_4	$\mathbf{0.126}_1$	0.364_3	0.272_2
	(0.324)	(**0.075**)	(0.471)	(0.064)
flare	2.047_4	$\mathbf{0.421}_1$	0.425_2	0.548_3
	(0.831)	(**0.023**)	(0.025)	(0.016)
german	0.786_4	0.536_2	0.550_3	$\mathbf{0.523}_1$
	(0.198)	(0.026)	(0.083)	(**0.025**)
glass	1.179_4	0.599_2	0.769_3	$\mathbf{0.598}_1$
	(0.774)	(0.053)	(0.461)	(**0.044**)
heart-s	0.568_4	$\mathbf{0.397}_1$	0.508_3	0.407_2
	(0.205)	(**0.079**)	(0.440)	(0.080)
hepatit	2.746_4	$\mathbf{0.365}_1$	0.467_3	0.405_2
	(1.776)	(**0.103**)	(0.442)	(0.072)
horse	2.503_4	$\mathbf{0.482}_1$	0.733_3	0.593_2
	(1.469)	(**0.052**)	(0.408)	(0.043)
ionosph	1.077_4	$\mathbf{0.293}_1$	0.425_3	0.319_2
	(0.393)	(**0.061**)	(0.370)	(0.068)
iris	$\mathbf{0.000}_1$	0.000_3	0.008_4	0.000_2
	(**0.002**)	(0.002)	(0.014)	(0.002)
landsat	0.630_4	$\mathbf{0.180}_1$	0.191_2	0.221_3
	(0.094)	(**0.016**)	(0.027)	(0.017)
letter	0.087_4	0.081_2	$\mathbf{0.077}_1$	0.085_3
	(0.006)	(0.003)	(**0.004**)	(0.003)
libras-	0.377_4	$\mathbf{0.124}_1$	0.253_3	0.132_2
	(0.379)	(**0.065**)	(0.286)	(0.054)
lung-ca	0.752_4	$\mathbf{0.164}_1$	0.171_3	0.169_2
	(0.991)	(**0.067**)	(0.070)	(0.068)
mfeat-k	0.123_4	$\mathbf{0.055}_1$	0.092_3	0.064_2
	(0.080)	(**0.019**)	(0.083)	(0.021)
mfeat-m	0.055_4	0.013_2	0.018_3	$\mathbf{0.013}_1$
	(0.072)	(0.017)	(0.041)	(**0.017**)
mfeat-z	1.189_4	$\mathbf{0.092}_1$	0.107_2	0.138_3
	(0.180)	(**0.022**)	(0.059)	(0.016)
mushroo	0.611_4	0.258_2	$\mathbf{0.228}_1$	0.269_3
	(0.087)	(0.015)	(**0.019**)	(0.019)
optdigi	11.967_4	$\mathbf{0.240}_1$	0.344_3	0.270_2
	(1.024)	(**0.010**)	(0.114)	(0.006)
page-bl	0.992_4	0.210_2	$\mathbf{0.194}_1$	0.261_3
	(0.147)	(0.012)	(**0.024**)	(0.011)
pendigi	0.328_4	$\mathbf{0.076}_1$	0.093_3	0.086_2
	(0.037)	(**0.006**)	(0.016)	(0.006)
scene-c	11.566_4	$\mathbf{0.452}_1$	0.582_3	0.463_2
	(0.720)	(**0.015**)	(0.096)	(0.013)
segment	1.256_4	$\mathbf{0.110}_1$	0.138_2	0.196_3
	(0.132)	(**0.022**)	(0.083)	(0.021)
shuttle	0.675_4	0.276_2	$\mathbf{0.206}_1$	0.322_3
	(0.025)	(0.003)	(**0.002**)	(0.003)
sonar	1.987_4	$\mathbf{0.538}_1$	0.643_3	0.592_2
	(0.941)	(**0.068**)	(0.397)	(0.062)
spambas	4.959_4	$\mathbf{0.359}_1$	0.701_3	0.403_2
	(0.462)	(**0.024**)	(0.136)	(0.015)
tic-tac	0.589_3	$\mathbf{0.579}_1$	0.589_4	0.582_2
	(0.016)	(**0.022**)	(0.085)	(0.022)
vehicle	0.426_4	$\mathbf{0.386}_1$	0.389_2	0.402_3
	(0.113)	(**0.040**)	(0.078)	(0.046)
vowel	$\mathbf{0.112}_1$	0.112_2	0.161_4	0.126_3
	(**0.024**)	(0.023)	(0.130)	(0.030)
wavefor	0.359_4	$\mathbf{0.291}_1$	0.307_3	0.306_2
	(0.030)	(**0.016**)	(0.032)	(0.017)
wdbc	0.511_4	$\mathbf{0.146}_1$	0.203_2	0.223_3
	(0.238)	(**0.040**)	(0.150)	(0.051)
wpbc	1.373_4	0.539_2	0.740_3	$\mathbf{0.537}_1$
	(0.593)	(0.047)	(0.576)	(**0.028**)
yeast	1.922_4	$\mathbf{0.550}_1$	0.557_2	0.567_3
	(1.945)	(**0.018**)	(0.048)	(0.028)
zoo	0.585_4	0.136_2	0.314_3	$\mathbf{0.136}_1$
	(1.050)	(0.198)	(0.628)	(**0.198**)
rank	3.80	1.32	2.63	2.24

Table 3: Brier score results for Naive Bayes (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.002e{-}14$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	0.302_4	0.216_2	$\mathbf{0.216}_1$	0.217_3
	(0.012)	(0.003)	(**0.004**)	(0.004)
autos	0.221_4	0.178_2	$\mathbf{0.178}_1$	0.189_3
	(0.054)	(0.018)	(**0.023**)	(0.016)
balance	0.077_4	0.034_2	$\mathbf{0.034}_1$	0.034_3
	(0.006)	(0.011)	(**0.011**)	(0.011)
car	0.134_4	0.132_2	$\mathbf{0.129}_1$	0.134_3
	(0.006)	(0.009)	(**0.009**)	(0.011)
clevela	0.138_4	$\mathbf{0.129}_1$	0.130_3	0.129_2
	(0.038)	(**0.027**)	(0.028)	(0.030)
credit-	0.162_4	$\mathbf{0.117}_1$	0.117_2	0.140_3
	(0.030)	(**0.018**)	(0.019)	(0.021)
dermato	0.011_3	$\mathbf{0.011}_1$	0.011_2	0.011_4
	(0.010)	(**0.009**)	(0.010)	(0.010)
diabete	0.179_4	$\mathbf{0.167}_1$	0.169_2	0.171_3
	(0.022)	(**0.014**)	(0.015)	(0.015)
ecoli	0.254_4	$\mathbf{0.032}_1$	0.034_2	0.080_3
	(0.050)	(**0.018**)	(0.020)	(0.023)
flare	0.462_4	$\mathbf{0.135}_1$	0.136_2	0.186_3
	(0.032)	(**0.008**)	(0.008)	(0.005)
german	0.197_4	0.178_3	0.174_2	$\mathbf{0.174}_1$
	(0.024)	(0.008)	(0.011)	(**0.010**)
glass	0.303_4	0.208_2	0.209_3	$\mathbf{0.208}_1$
	(0.069)	(0.017)	(0.024)	(**0.016**)
heart-s	0.130_4	$\mathbf{0.122}_1$	0.123_3	0.123_2
	(0.039)	(**0.028**)	(0.030)	(0.031)
hepatit	0.308_4	$\mathbf{0.112}_1$	0.114_2	0.128_3
	(0.097)	(**0.033**)	(0.042)	(0.021)
horse	0.384_4	0.152_2	$\mathbf{0.149}_1$	0.202_3
	(0.104)	(0.020)	(**0.023**)	(0.019)
ionosph	0.099_4	0.087_2	$\mathbf{0.085}_1$	0.090_3
	(0.030)	(0.021)	(**0.023**)	(0.024)
iris	0.000_3	$\mathbf{0.000}_1$	0.001_4	0.000_2
	(0.001)	(**0.000**)	(0.002)	(0.001)
landsat	0.060_4	0.050_2	$\mathbf{0.050}_1$	0.057_3
	(0.006)	(0.005)	(**0.005**)	(0.006)
letter	0.021_3	0.021_2	$\mathbf{0.020}_1$	0.021_4
	(0.001)	(0.001)	(**0.001**)	(0.001)
libras-	0.038_4	0.029_2	$\mathbf{0.029}_1$	0.034_3
	(0.021)	(0.014)	(**0.015**)	(0.014)
lung-ca	$\mathbf{0.039}_1$	0.039_2	0.041_4	0.041_3
	(**0.023**)	(0.021)	(0.021)	(0.021)
mfeat-k	0.013_3	$\mathbf{0.013}_1$	0.013_2	0.014_4
	(0.005)	(**0.004**)	(0.004)	(0.005)
mfeat-m	0.002_3	0.001_2	0.002_4	$\mathbf{0.001}_1$
	(0.002)	(0.002)	(0.002)	(**0.002**)
mfeat-z	0.071_4	0.024_2	$\mathbf{0.022}_1$	0.043_3
	(0.008)	(0.005)	(**0.006**)	(0.004)
mushroo	0.078_4	0.074_3	$\mathbf{0.062}_1$	0.071_2
	(0.008)	(0.005)	(**0.004**)	(0.006)
optdigi	0.425_4	$\mathbf{0.076}_1$	0.084_3	0.083_2
	(0.020)	(**0.002**)	(0.006)	(0.001)
page-bl	0.091_4	0.062_2	$\mathbf{0.057}_1$	0.072_3
	(0.009)	(0.004)	(**0.003**)	(0.004)
pendigi	0.027_4	$\mathbf{0.022}_1$	0.026_3	0.023_2
	(0.002)	(**0.002**)	(0.002)	(0.002)
scene-c	0.382_4	$\mathbf{0.148}_1$	0.160_3	0.152_2
	(0.021)	(**0.004**)	(0.008)	(0.004)
segment	0.148_4	0.030_2	$\mathbf{0.029}_1$	0.066_3
	(0.013)	(0.006)	(**0.006**)	(0.004)
shuttle	0.098_4	0.083_2	$\mathbf{0.069}_1$	0.090_3
	(0.001)	(0.001)	(**0.001**)	(0.001)
sonar	0.265_4	0.182_2	$\mathbf{0.177}_1$	0.202_3
	(0.063)	(0.029)	(**0.031**)	(0.028)
spambas	0.177_4	$\mathbf{0.111}_1$	0.157_3	0.130_2
	(0.011)	(**0.009**)	(0.017)	(0.006)
tic-tac	0.201_4	0.196_2	$\mathbf{0.194}_1$	0.198_3
	(0.007)	(0.010)	(**0.011**)	(0.010)
vehicle	0.131_4	0.123_2	$\mathbf{0.120}_1$	0.126_3
	(0.023)	(0.016)	(**0.015**)	(0.018)
vowel	0.035_4	$\mathbf{0.034}_1$	0.034_2	0.035_3
	(0.008)	(**0.008**)	(0.008)	(0.009)
wavefor	0.108_4	$\mathbf{0.094}_1$	0.094_2	0.096_3
	(0.008)	(**0.006**)	(0.006)	(0.006)
wdbc	0.059_4	0.044_2	$\mathbf{0.044}_1$	0.057_3
	(0.018)	(0.013)	(**0.013**)	(0.017)
wpbc	0.260_4	0.177_2	0.180_3	$\mathbf{0.177}_1$
	(0.054)	(0.017)	(0.028)	(**0.011**)
yeast	0.393_4	0.187_2	$\mathbf{0.184}_1$	0.193_3
	(0.140)	(0.006)	(**0.008**)	(0.011)
zoo	0.019_4	0.018_3	0.018_2	$\mathbf{0.018}_1$
	(0.030)	(0.028)	(0.028)	(**0.028**)
rank	3.80	1.68	1.88	2.63

Table 4: Accuracy results for Naive Bayes in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $3.201e{-}05$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	61.851_4	63.379_2	$\mathbf{63.393}_1$	63.043_3
	(1.430)	(0.917)	(**1.045**)	(1.314)
autos	68.503_4	$\mathbf{72.031}_1$	72.022_2	70.319_3
	(8.098)	(**6.241**)	(6.661)	(5.107)
balance	95.346_4	$\mathbf{95.986}_1$	95.794_3	95.986_2
	(1.682)	(**2.023**)	(2.146)	(2.063)
car	$\mathbf{78.727}_1$	78.513_4	78.704_2	78.641_3
	(**2.223**)	(2.269)	(2.212)	(2.243)
clevela	$\mathbf{83.641}_1$	83.441_3	83.239_4	83.540_2
	(**4.615**)	(4.642)	(4.250)	(4.703)
credit-	80.706_4	84.488_2	$\mathbf{84.533}_1$	81.458_3
	(3.500)	(2.891)	(**2.972**)	(3.341)
dermato	$98.825_{2.5}$	$98.825_{2.5}$	$98.825_{2.5}$	$98.825_{2.5}$
	(1.181)	(1.181)	(1.181)	(1.181)
diabete	$\mathbf{75.389}_1$	74.830_4	75.038_3	75.273_2
	(**3.589**)	(3.398)	(3.398)	(3.358)
ecoli	65.604_4	$\mathbf{95.994}_1$	95.753_2	93.703_3
	(8.741)	(**2.595**)	(2.879)	(3.113)
flare	50.918_4	$\mathbf{82.289}_1$	81.955_2	71.489_3
	(3.326)	(**1.133**)	(1.311)	(0.140)
german	72.930_3	72.510_4	$\mathbf{74.020}_1$	73.990_2
	(2.770)	(1.973)	(**2.243**)	(2.327)
glass	58.673_4	$\mathbf{63.880}_1$	63.824_2	63.521_3
	(5.251)	(**3.605**)	(6.312)	(3.769)
heart-s	84.296_3	$\mathbf{84.481}_1$	84.333_2	84.148_4
	(5.051)	(**4.731**)	(5.022)	(4.630)
hepatit	60.419_4	84.343_2	$\mathbf{84.416}_1$	79.565_3
	(10.779)	(5.538)	(**5.792**)	(1.446)
horse	50.253_4	80.428_2	$\mathbf{81.092}_1$	67.407_3
	(10.602)	(3.822)	(**3.515**)	(6.242)
ionosph	88.687_4	89.545_2	$\mathbf{89.800}_1$	89.344_3
	(3.396)	(3.328)	(**3.685**)	(3.365)
iris	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
landsat	93.529_4	93.812_2	$\mathbf{93.826}_1$	93.616_3
	(0.674)	(0.654)	(**0.666**)	(0.694)
letter	97.264_2	97.246_3	$\mathbf{97.363}_1$	97.117_4
	(0.109)	(0.201)	(**0.174**)	(0.186)
libras-	95.160_4	96.293_2	$\mathbf{96.293}_1$	95.660_3
	(2.659)	(2.038)	(**1.992**)	(2.390)
lung-ca	$94.779_{1.5}$	$94.779_{1.5}$	93.732_4	93.942_3
	(3.512)	(3.512)	(3.669)	(3.566)
mfeat-k	98.395_2	98.370_4	98.385_3	$\mathbf{98.400}_1$
	(0.600)	(0.579)	(0.609)	(**0.598**)
mfeat-m	$99.850_{2.5}$	$99.850_{2.5}$	$99.850_{2.5}$	$99.850_{2.5}$
	(0.196)	(0.196)	(0.196)	(0.196)
mfeat-z	92.640_4	97.045_2	$\mathbf{97.350}_1$	93.165_3
	(0.918)	(0.806)	(**0.751**)	(0.815)
mushroo	90.971_3	90.314_4	$\mathbf{92.613}_1$	91.723_2
	(1.057)	(0.703)	(**0.532**)	(0.576)
optdigi	54.229_4	$89.822_{1.5}$	88.465_3	$89.822_{1.5}$
	(2.466)	(0.021)	(1.854)	(0.021)
page-bl	89.880_4	90.926_3	$\mathbf{92.147}_1$	91.139_2
	(1.014)	(1.027)	(**0.463**)	(0.937)
pendigi	$\mathbf{97.233}_1$	97.187_3	96.063_4	97.216_2
	(**0.235**)	(0.254)	(0.546)	(0.247)
scene-c	60.597_4	$77.856_{1.5}$	73.972_3	$77.856_{1.5}$
	(2.161)	(0.070)	(3.862)	(0.070)
segment	83.753_4	96.095_2	$\mathbf{96.429}_1$	90.039_3
	(1.533)	(0.847)	(**0.753**)	(1.360)
shuttle	89.458_4	89.489_3	$\mathbf{89.916}_1$	89.521_2
	(0.172)	(0.146)	(**0.162**)	(0.199)
sonar	68.125_4	71.225_2	$\mathbf{73.008}_1$	69.590_3
	(7.382)	(7.470)	(**7.082**)	(7.854)
spambas	81.387_3	$\mathbf{84.210}_1$	66.457_4	82.317_2
	(1.272)	(**1.116**)	(7.906)	(1.151)
tic-tac	71.314_4	71.377_3	$\mathbf{71.543}_1$	71.471_2
	(1.837)	(2.679)	(**2.415**)	(2.789)
vehicle	80.980_4	82.035_2	$\mathbf{82.223}_1$	81.572_3
	(3.527)	(2.861)	(**2.927**)	(2.837)
vowel	95.010_4	95.323_3	95.354_2	$\mathbf{95.364}_1$
	(1.305)	(1.376)	(1.279)	(**1.317**)
wavefor	85.590_4	$\mathbf{86.388}_1$	86.288_3	86.346_2
	(1.174)	(**1.152**)	(1.174)	(1.136)
wdbc	$93.356_{1.5}$	93.338_3	93.249_4	$93.356_{1.5}$
	(2.005)	(1.959)	(1.979)	(2.126)
wpbc	67.516_4	75.677_3	76.189_2	$\mathbf{76.296}_1$
	(6.215)	(2.751)	(4.307)	(**0.710**)
yeast	49.260_4	69.117_2	$\mathbf{69.606}_1$	68.658_3
	(10.057)	(1.362)	(**2.139**)	(0.488)
zoo	$98.258_{2.5}$	$98.258_{2.5}$	$98.258_{2.5}$	$98.258_{2.5}$
	(3.024)	(3.024)	(3.024)	(3.024)
rank	3.24	2.28	2.00	2.48

Table 5: Log-loss results for Naive Bayes (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.202e{-}20$, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	1.170_5	$\mathbf{0.621}_1$	0.631_4	0.622_2	0.623_3
	(0.077)	(**0.008**)	(0.008)	(0.008)	(0.008)
autos	0.810_5	$\mathbf{0.529}_1$	0.545_2	0.549_3	0.558_4
	(0.305)	(**0.043**)	(0.052)	(0.041)	(0.037)
balance	0.284_5	$\mathbf{0.117}_1$	0.122_4	0.119_3	0.118_2
	(0.013)	(**0.033**)	(0.034)	(0.032)	(0.032)
car	0.396_4	0.382_2	0.389_3	$\mathbf{0.382}_1$	0.396_5
	(0.015)	(0.022)	(0.019)	(**0.023**)	(0.026)
clevela	0.631_5	0.415_3	$\mathbf{0.412}_1$	0.414_2	0.424_4
	(0.220)	(0.074)	(**0.073**)	(0.073)	(0.076)
credit-	1.043_5	$\mathbf{0.390}_1$	0.416_3	0.409_2	0.448_4
	(0.332)	(**0.051**)	(0.051)	(0.059)	(0.055)
dermato	0.282_5	0.060_2	0.060_3	$\mathbf{0.057}_1$	0.062_4
	(0.340)	(0.047)	(0.046)	(**0.042**)	(0.047)
diabete	0.612_5	$\mathbf{0.502}_1$	0.502_2	0.528_4	0.516_3
	(0.116)	(**0.032**)	(0.031)	(0.042)	(0.034)
ecoli	1.376_5	0.126_2	0.445_4	$\mathbf{0.125}_1$	0.272_3
	(0.324)	(0.075)	(0.065)	(**0.076**)	(0.064)
flare	2.047_5	$\mathbf{0.421}_1$	0.582_4	0.439_2	0.548_3
	(0.831)	(**0.023**)	(0.023)	(0.029)	(0.016)
german	0.786_5	0.536_2	0.564_4	0.561_3	$\mathbf{0.523}_1$
	(0.198)	(0.026)	(0.045)	(0.023)	(**0.025**)
glass	1.179_5	0.599_3	0.609_4	$\mathbf{0.591}_1$	0.598_2
	(0.774)	(0.053)	(0.051)	(**0.052**)	(0.044)
heart-s	0.568_5	0.397_3	0.395_2	$\mathbf{0.395}_1$	0.407_4
	(0.205)	(0.079)	(0.077)	(**0.077**)	(0.080)
hepatit	2.746_5	$\mathbf{0.365}_1$	0.503_4	0.377_2	0.405_3
	(1.776)	(**0.103**)	(0.083)	(0.090)	(0.072)
horse	2.503_5	$\mathbf{0.482}_1$	0.629_4	0.520_2	0.593_3
	(1.469)	(**0.052**)	(0.056)	(0.046)	(0.043)
ionosph	1.077_5	0.293_2	0.294_3	$\mathbf{0.291}_1$	0.319_4
	(0.393)	(0.061)	(0.058)	(**0.060**)	(0.068)
iris	$\mathbf{0.000}_1$	0.000_3	0.000_4	0.000_5	0.000_2
	(**0.002**)	(0.002)	(0.002)	(0.002)	(0.002)
landsat	0.630_5	$\mathbf{0.180}_1$	0.184_3	0.183_2	0.221_4
	(0.094)	(**0.016**)	(0.016)	(0.016)	(0.017)
letter	0.087_5	$\mathbf{0.081}_1$	0.084_3	0.083_2	0.085_4
	(0.006)	(**0.003**)	(0.004)	(0.004)	(0.003)
libras-	0.377_5	$\mathbf{0.124}_1$	0.140_4	0.126_2	0.132_3
	(0.379)	(**0.065**)	(0.060)	(0.057)	(0.054)
lung-ca	0.752_5	0.164_2	0.164_3	$\mathbf{0.159}_1$	0.169_4
	(0.991)	(0.067)	(0.067)	(**0.067**)	(0.068)
mfeat-k	0.123_5	$\mathbf{0.055}_1$	0.056_2	0.060_3	0.064_4
	(0.080)	(**0.019**)	(0.019)	(0.022)	(0.021)
mfeat-m	0.055_5	0.013_3	0.013_4	0.013_2	$\mathbf{0.013}_1$
	(0.072)	(0.017)	(0.017)	(0.017)	(**0.017**)
mfeat-z	1.189_5	$\mathbf{0.092}_1$	0.135_3	0.097_2	0.138_4
	(0.180)	(**0.022**)	(0.016)	(0.021)	(0.016)
mushroo	0.611_5	$\mathbf{0.258}_1$	0.318_4	0.309_3	0.269_2
	(0.087)	(**0.015**)	(0.017)	(0.022)	(0.019)
optdigi	11.967_5	$\mathbf{0.240}_1$	0.313_4	0.249_2	0.270_3
	(1.024)	(**0.010**)	(0.006)	(0.011)	(0.006)
page-bl	0.992_5	$\mathbf{0.210}_1$	0.210_2	0.227_3	0.261_4
	(0.147)	(**0.012**)	(0.012)	(0.011)	(0.011)
pendigi	0.328_5	$\mathbf{0.076}_1$	0.077_3	0.077_2	0.086_4
	(0.037)	(**0.006**)	(0.006)	(0.006)	(0.006)
scene-c	11.566_5	$\mathbf{0.452}_1$	0.480_4	0.453_2	0.463_3
	(0.720)	(**0.015**)	(0.014)	(0.015)	(0.013)
segment	1.256_5	$\mathbf{0.110}_1$	0.239_4	0.113_2	0.196_3
	(0.132)	(**0.022**)	(0.014)	(0.022)	(0.021)
shuttle	0.675_5	$\mathbf{0.276}_1$	0.277_2	0.285_3	0.322_4
	(0.025)	(**0.003**)	(0.003)	(0.003)	(0.003)
sonar	1.987_5	$\mathbf{0.538}_1$	0.563_3	0.562_2	0.592_4
	(0.941)	(**0.068**)	(0.068)	(0.070)	(0.062)
spambas	4.959_5	$\mathbf{0.359}_1$	0.391_3	0.373_2	0.403_4
	(0.462)	(**0.024**)	(0.017)	(0.019)	(0.015)
tic-tac	0.589_5	$\mathbf{0.579}_1$	0.587_4	0.583_3	0.582_2
	(0.016)	(**0.022**)	(0.017)	(0.022)	(0.022)
vehicle	0.426_5	$\mathbf{0.386}_1$	0.393_2	0.412_4	0.402_3
	(0.113)	(**0.040**)	(0.041)	(0.041)	(0.046)
vowel	0.112_2	0.112_3	0.113_4	$\mathbf{0.111}_1$	0.126_5
	(0.024)	(0.023)	(0.023)	(**0.023**)	(0.030)
wavefor	0.359_5	$\mathbf{0.291}_1$	0.302_3	0.292_2	0.306_4
	(0.030)	(**0.016**)	(0.016)	(0.016)	(0.017)
wdbc	0.511_5	0.146_3	0.145_2	$\mathbf{0.144}_1$	0.223_4
	(0.238)	(0.040)	(0.039)	(**0.038**)	(0.051)
wpbc	1.373_5	0.539_3	0.592_4	$\mathbf{0.533}_1$	0.537_2
	(0.593)	(0.047)	(0.064)	(**0.041**)	(0.028)
yeast	1.922_5	$\mathbf{0.550}_1$	0.610_4	0.576_3	0.567_2
	(1.945)	(**0.018**)	(0.029)	(0.022)	(0.028)
zoo	0.585_5	0.136_2	0.136_3	0.136_4	$\mathbf{0.136}_1$
	(1.050)	(0.198)	(0.198)	(0.198)	(**0.198**)
rank	4.80	1.56	3.22	2.20	3.22

Table 6: Brier score results for Naive Bayes (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $2.368e{-}18$, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	0.302_5	$\mathbf{0.216}_1$	0.220_4	0.216_2	0.217_3
	(0.012)	(**0.003**)	(0.003)	(0.003)	(0.004)
autos	0.221_5	$\mathbf{0.178}_1$	0.184_2	0.186_3	0.189_4
	(0.054)	(**0.018**)	(0.020)	(0.016)	(0.016)
balance	0.077_5	$\mathbf{0.034}_1$	0.035_4	0.034_3	0.034_2
	(0.006)	(**0.011**)	(0.012)	(0.011)	(0.011)
car	0.134_5	0.132_2	0.134_3	$\mathbf{0.132}_1$	0.134_4
	(0.006)	(0.009)	(0.008)	(**0.010**)	(0.011)
clevela	0.138_5	0.129_3	$\mathbf{0.128}_1$	0.128_2	0.129_4
	(0.038)	(0.027)	(**0.027**)	(0.027)	(0.030)
credit-	0.162_5	$\mathbf{0.117}_1$	0.131_3	0.123_2	0.140_4
	(0.030)	(**0.018**)	(0.018)	(0.018)	(0.021)
dermato	0.011_3	0.011_2	0.011_4	$\mathbf{0.011}_1$	0.011_5
	(0.010)	(0.009)	(0.009)	(**0.009**)	(0.010)
diabete	0.179_5	0.167_2	$\mathbf{0.167}_1$	0.171_3	0.171_4
	(0.022)	(0.014)	(**0.013**)	(0.012)	(0.015)
ecoli	0.254_5	0.032_2	0.146_4	$\mathbf{0.032}_1$	0.080_3
	(0.050)	(0.018)	(0.023)	(**0.018**)	(0.023)
flare	0.462_5	$\mathbf{0.135}_1$	0.204_4	0.137_2	0.186_3
	(0.032)	(**0.008**)	(0.008)	(0.007)	(0.005)
german	0.197_5	0.178_2	0.184_3	0.186_4	$\mathbf{0.174}_1$
	(0.024)	(0.008)	(0.011)	(0.007)	(**0.010**)
glass	0.303_5	0.208_3	0.213_4	$\mathbf{0.204}_1$	0.208_2
	(0.069)	(0.017)	(0.016)	(**0.014**)	(0.016)
heart-s	0.130_5	0.122_3	0.122_2	$\mathbf{0.122}_1$	0.123_4
	(0.039)	(0.028)	(0.027)	(**0.027**)	(0.031)
hepatit	0.308_5	$\mathbf{0.112}_1$	0.168_4	0.114_2	0.128_3
	(0.097)	(**0.033**)	(0.028)	(0.027)	(0.021)
horse	0.384_5	$\mathbf{0.152}_1$	0.219_4	0.161_2	0.202_3
	(0.104)	(**0.020**)	(0.025)	(0.016)	(0.019)
ionosph	0.099_5	0.087_2	0.088_3	$\mathbf{0.087}_1$	0.090_4
	(0.030)	(0.021)	(0.021)	(**0.021**)	(0.024)
iris	0.000_3	$\mathbf{0.000}_1$	0.000_4	0.000_5	0.000_2
	(0.001)	(**0.000**)	(0.001)	(0.001)	(0.001)
landsat	0.060_5	$\mathbf{0.050}_1$	0.051_3	0.050_2	0.057_4
	(0.006)	(**0.005**)	(0.005)	(0.005)	(0.006)
letter	0.021_3	0.021_2	0.021_4	$\mathbf{0.021}_1$	0.021_5
	(0.001)	(0.001)	(0.001)	(**0.001**)	(0.001)
libras-	0.038_5	$\mathbf{0.029}_1$	0.035_4	0.031_2	0.034_3
	(0.021)	(**0.014**)	(0.015)	(0.013)	(0.014)
lung-ca	0.039_2	0.039_3	0.039_4	$\mathbf{0.037}_1$	0.041_5
	(0.023)	(0.021)	(0.021)	(**0.021**)	(0.021)
mfeat-k	0.013_4	0.013_2	$\mathbf{0.013}_1$	0.013_3	0.014_5
	(0.005)	(0.004)	(**0.004**)	(0.004)	(0.005)
mfeat-m	0.002_5	0.001_4	0.001_3	0.001_2	$\mathbf{0.001}_1$
	(0.002)	(0.002)	(0.002)	(0.002)	(**0.002**)
mfeat-z	0.071_5	$\mathbf{0.024}_1$	0.042_3	0.025_2	0.043_4
	(0.008)	(**0.005**)	(0.004)	(0.005)	(0.004)
mushroo	0.078_3	0.074_2	0.096_5	0.079_4	$\mathbf{0.071}_1$
	(0.008)	(0.005)	(0.005)	(0.004)	(**0.006**)
optdigi	0.425_5	$\mathbf{0.076}_1$	0.096_4	0.077_2	0.083_3
	(0.020)	(**0.002**)	(0.002)	(0.002)	(0.001)
page-bl	0.091_5	$\mathbf{0.062}_1$	0.063_2	0.064_3	0.072_4
	(0.009)	(**0.004**)	(0.004)	(0.004)	(0.004)
pendigi	0.027_5	$\mathbf{0.022}_1$	0.022_2	0.023_3	0.023_4
	(0.002)	(**0.002**)	(0.002)	(0.002)	(0.002)
scene-c	0.382_5	$\mathbf{0.148}_1$	0.160_4	0.148_2	0.152_3
	(0.021)	(**0.004**)	(0.004)	(0.004)	(0.004)
segment	0.148_5	$\mathbf{0.030}_1$	0.082_4	0.031_2	0.066_3
	(0.013)	(**0.006**)	(0.005)	(0.005)	(0.004)
shuttle	0.098_5	0.083_3	0.083_2	$\mathbf{0.083}_1$	0.090_4
	(0.001)	(0.001)	(0.001)	(**0.001**)	(0.001)
sonar	0.265_5	$\mathbf{0.182}_1$	0.191_3	0.187_2	0.202_4
	(0.063)	(**0.029**)	(0.029)	(0.027)	(0.028)
spambas	0.177_5	$\mathbf{0.111}_1$	0.126_3	0.118_2	0.130_4
	(0.011)	(**0.009**)	(0.006)	(0.007)	(0.006)
tic-tac	0.201_5	$\mathbf{0.196}_1$	0.200_4	0.199_3	0.198_2
	(0.007)	(**0.010**)	(0.007)	(0.009)	(0.010)
vehicle	0.131_5	$\mathbf{0.123}_1$	0.126_2	0.130_4	0.126_3
	(0.023)	(**0.016**)	(0.016)	(0.014)	(0.018)
vowel	0.035_4	$\mathbf{0.034}_1$	0.035_5	0.034_2	0.035_3
	(0.008)	(**0.008**)	(0.008)	(0.007)	(0.009)
wavefor	0.108_5	$\mathbf{0.094}_1$	0.099_4	0.094_2	0.096_3
	(0.008)	(**0.006**)	(0.006)	(0.006)	(0.006)
wdbc	0.059_5	0.044_3	$\mathbf{0.043}_1$	0.044_2	0.057_4
	(0.018)	(0.013)	(**0.012**)	(0.012)	(0.017)
wpbc	0.260_5	0.177_3	0.199_4	$\mathbf{0.174}_1$	0.177_2
	(0.054)	(0.017)	(0.018)	(**0.013**)	(0.011)
yeast	0.393_5	$\mathbf{0.187}_1$	0.211_4	0.194_3	0.193_2
	(0.140)	(**0.006**)	(0.013)	(0.004)	(0.011)
zoo	0.019_5	0.018_3	0.018_4	$\mathbf{0.018}_1$	0.018_2
	(0.030)	(0.028)	(0.028)	(**0.028**)	(0.028)
rank	4.68	1.68	3.24	2.15	3.24

Table 7: Accuracy results for Naive Bayes in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $3.844e{-}08$, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	61.851_4	63.379_2	61.343_5	$\mathbf{63.415}_1$	63.043_3
	(1.430)	(0.917)	(1.398)	(**0.949**)	(1.314)
autos	68.503_5	$\mathbf{72.031}_1$	70.429_2	69.915_4	70.319_3
	(8.098)	(**6.241**)	(6.128)	(5.459)	(5.107)
balance	95.346_5	95.986_2	95.460_4	$\mathbf{96.001}_1$	95.986_3
	(1.682)	(2.023)	(1.754)	(**2.054**)	(2.063)
car	78.727_2	78.513_5	$\mathbf{78.780}_1$	78.640_4	78.641_3
	(2.223)	(2.269)	(**2.253**)	(2.371)	(2.243)
clevela	$\mathbf{83.641}_1$	83.441_4	83.372_5	83.474_3	83.540_2
	(**4.615**)	(4.642)	(4.647)	(4.412)	(4.703)
credit-	80.706_4	$\mathbf{84.488}_1$	80.645_5	83.143_2	81.458_3
	(3.500)	(**2.891**)	(3.589)	(3.069)	(3.341)
dermato	98.825_3	98.825_3	98.825_3	98.825_3	98.825_3
	(1.181)	(1.181)	(1.181)	(1.181)	(1.181)
diabete	75.389_2	74.830_4	$\mathbf{75.454}_1$	74.687_5	75.273_3
	(3.589)	(3.398)	(**3.580**)	(2.929)	(3.358)
ecoli	65.604_5	$95.994_{1.5}$	77.200_4	$95.994_{1.5}$	93.703_3
	(8.741)	(2.595)	(4.090)	(2.595)	(3.113)
flare	50.918_5	$82.289_{1.5}$	52.459_4	$82.289_{1.5}$	71.489_3
	(3.326)	(1.133)	(3.788)	(1.133)	(0.140)
german	72.930_3	72.510_4	73.430_2	70.980_5	$\mathbf{73.990}_1$
	(2.770)	(1.973)	(2.339)	(1.297)	(**2.327**)
glass	58.673_5	63.880_2	59.906_4	$\mathbf{64.570}_1$	63.521_3
	(5.251)	(3.605)	(6.696)	(**3.347**)	(3.769)
heart-s	84.296_3	84.481_2	84.000_5	$\mathbf{84.593}_1$	84.148_4
	(5.051)	(4.731)	(5.194)	(**4.402**)	(4.630)
hepatit	60.419_5	84.343_2	72.221_4	$\mathbf{84.359}_1$	79.565_3
	(10.779)	(5.538)	(8.633)	(**5.035**)	(1.446)
horse	50.253_5	80.428_2	63.798_4	$\mathbf{80.661}_1$	67.407_3
	(10.602)	(3.822)	(12.957)	(**3.420**)	(6.242)
ionosph	88.687_5	89.545_2	88.829_4	$\mathbf{90.001}_1$	89.344_3
	(3.396)	(3.328)	(3.374)	(**3.451**)	(3.365)
iris	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
landsat	93.529_5	$\mathbf{93.812}_1$	93.532_4	93.742_2	93.616_3
	(0.674)	(**0.654**)	(0.669)	(0.638)	(0.694)
letter	97.264_2	97.246_3	$\mathbf{97.279}_1$	97.041_5	97.117_4
	(0.109)	(0.201)	(**0.107**)	(0.190)	(0.186)
libras-	95.160_5	$\mathbf{96.293}_1$	95.933_3	96.153_2	95.660_4
	(2.659)	(**2.038**)	(2.646)	(2.046)	(2.390)
lung-ca	94.779_3	94.779_3	94.779_3	$\mathbf{94.884}_1$	93.942_5
	(3.512)	(3.512)	(3.512)	(**3.428**)	(3.566)
mfeat-k	$98.395_{3.5}$	98.370_5	$98.395_{3.5}$	$98.400_{1.5}$	$98.400_{1.5}$
	(0.600)	(0.579)	(0.608)	(0.612)	(0.598)
mfeat-m	99.850_3	99.850_3	99.850_3	99.850_3	99.850_3
	(0.196)	(0.196)	(0.196)	(0.196)	(0.196)
mfeat-z	92.640_5	$\mathbf{97.045}_1$	92.780_4	96.810_2	93.165_3
	(0.918)	(**0.806**)	(0.893)	(0.815)	(0.815)
mushroo	90.971_3	90.314_5	90.944_4	91.400_2	$\mathbf{91.723}_1$
	(1.057)	(0.703)	(1.193)	(0.633)	(**0.576**)
optdigi	54.229_5	89.822_2	89.815_4	89.822_2	89.822_2
	(2.466)	(0.021)	(0.032)	(0.021)	(0.021)
page-bl	89.880_5	90.926_3	89.913_4	$\mathbf{92.103}_1$	91.139_2
	(1.014)	(1.027)	(1.016)	(**0.498**)	(0.937)
pendigi	97.233_2	97.187_4	$\mathbf{97.234}_1$	97.163_5	97.216_3
	(0.235)	(0.254)	(**0.233**)	(0.274)	(0.247)
scene-c	60.597_5	77.856_2	77.798_4	77.856_2	77.856_2
	(2.161)	(0.070)	(0.126)	(0.070)	(0.070)
segment	83.753_5	$\mathbf{96.095}_1$	84.126_4	96.013_2	90.039_3
	(1.533)	(**0.847**)	(1.510)	(0.873)	(1.360)
shuttle	89.458_5	89.489_3	89.464_4	$\mathbf{89.713}_1$	89.521_2
	(0.172)	(0.146)	(0.171)	(**0.165**)	(0.199)
sonar	68.125_5	$\mathbf{71.225}_1$	68.744_4	70.078_2	69.590_3
	(7.382)	(**7.470**)	(6.926)	(7.558)	(7.854)
spambas	81.387_5	$\mathbf{84.210}_1$	82.067_4	83.408_2	82.317_3
	(1.272)	(**1.116**)	(1.147)	(1.099)	(1.151)
tic-tac	71.314_5	71.377_4	71.419_3	71.460_2	$\mathbf{71.471}_1$
	(1.837)	(2.679)	(1.812)	(2.748)	(**2.789**)
vehicle	80.980_5	$\mathbf{82.035}_1$	81.716_2	81.435_4	81.572_3
	(3.527)	(**2.861**)	(3.825)	(2.771)	(2.837)
vowel	95.010_5	$95.323_{2.5}$	95.040_4	$95.323_{2.5}$	$\mathbf{95.364}_1$
	(1.305)	(1.376)	(1.325)	(1.376)	(**1.317**)
wavefor	85.590_5	$\mathbf{86.388}_1$	85.642_4	86.368_2	86.346_3
	(1.174)	(**1.152**)	(1.181)	(1.099)	(1.136)
wdbc	$93.356_{3.5}$	93.338_5	93.373_2	$\mathbf{93.463}_1$	$93.356_{3.5}$
	(2.005)	(1.959)	(2.091)	(**2.047**)	(2.126)
wpbc	67.516_5	75.677_3	71.709_4	76.094_2	$\mathbf{76.296}_1$
	(6.215)	(2.751)	(5.876)	(1.629)	(**0.710**)
yeast	49.260_5	$\mathbf{69.117}_1$	68.766_2	68.470_4	68.658_3
	(10.057)	(**1.362**)	(0.518)	(0.665)	(0.488)
zoo	98.258_3	98.258_3	98.258_3	98.258_3	98.258_3
	(3.024)	(3.024)	(3.024)	(3.024)	(3.024)
rank	4.10	2.48	3.38	2.32	2.73

Table 8: Log-loss results for Ada-O (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.008e{-}12$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	0.612_2	$\mathbf{0.612}_1$	0.626_4	0.614_3
	(0.011)	(**0.010**)	(0.035)	(0.009)
autos	0.427_4	$\mathbf{0.283}_1$	0.416_3	0.287_2
	(0.359)	(**0.073**)	(0.408)	(0.066)
balance	0.049_4	$\mathbf{0.037}_1$	0.041_3	0.038_2
	(0.013)	(**0.018**)	(0.048)	(0.019)
car	0.114_2	$\mathbf{0.109}_1$	0.122_4	0.119_3
	(0.014)	(**0.018**)	(0.056)	(0.021)
clevela	0.496_3	$\mathbf{0.417}_1$	0.523_4	0.429_2
	(0.150)	(**0.051**)	(0.248)	(0.052)
credit-	0.360_3	0.341_2	0.444_4	$\mathbf{0.338}_1$
	(0.080)	(0.045)	(0.229)	(**0.041**)
dermato	0.036_4	$\mathbf{0.019}_1$	0.035_3	0.021_2
	(0.108)	(**0.035**)	(0.107)	(0.038)
diabete	0.506_3	$\mathbf{0.484}_1$	0.525_4	0.495_2
	(0.050)	(**0.028**)	(0.099)	(0.028)
ecoli	0.326_4	$\mathbf{0.145}_1$	0.224_3	0.159_2
	(0.372)	(**0.072**)	(0.269)	(0.073)
flare	$\mathbf{0.404}_1$	0.404_2	0.421_4	0.408_3
	(**0.023**)	(0.023)	(0.048)	(0.024)
german	0.511_3	0.506_2	0.538_4	$\mathbf{0.506}_1$
	(0.038)	(0.023)	(0.102)	(**0.021**)
glass	0.696_4	0.489_2	0.560_3	$\mathbf{0.485}_1$
	(0.266)	(0.059)	(0.269)	(**0.054**)
heart-s	0.570_4	$\mathbf{0.427}_1$	0.557_3	0.443_2
	(0.212)	(**0.062**)	(0.302)	(0.060)
hepatit	0.805_4	$\mathbf{0.392}_1$	0.431_3	0.411_2
	(0.507)	(**0.056**)	(0.232)	(0.061)
horse	0.642_4	$\mathbf{0.413}_1$	0.524_3	0.418_2
	(0.232)	(**0.053**)	(0.261)	(0.047)
ionosph	0.381_4	$\mathbf{0.203}_1$	0.296_3	0.227_2
	(0.203)	(**0.044**)	(0.212)	(0.045)
iris	0.127_4	0.000_2	$\mathbf{0.000}_1$	0.000_3
	(0.000)	(0.000)	(**0.000**)	(0.000)
landsat	0.041_2	$\mathbf{0.040}_1$	0.043_3	0.053_4
	(0.008)	(**0.006**)	(0.011)	(0.009)
letter	0.037_3	0.035_2	$\mathbf{0.035}_1$	0.044_4
	(0.002)	(0.002)	(**0.003**)	(0.003)
libras-	0.589_4	$\mathbf{0.100}_1$	0.184_3	0.112_2
	(0.502)	(**0.054**)	(0.261)	(0.053)
lung-ca	0.193_4	$\mathbf{0.139}_1$	0.189_3	0.139_2
	(0.051)	(**0.178**)	(0.534)	(0.184)
mfeat-k	0.062_4	$\mathbf{0.027}_1$	0.048_3	0.032_2
	(0.057)	(**0.013**)	(0.047)	(0.013)
mfeat-m	0.040_4	$\mathbf{0.014}_1$	0.026_3	0.014_2
	(0.060)	(**0.017**)	(0.043)	(0.016)
mfeat-z	0.095_4	$\mathbf{0.032}_1$	0.063_3	0.039_2
	(0.075)	(**0.014**)	(0.044)	(0.017)
mushroo	0.000_4	0.000_3	0.000_2	$\mathbf{0.000}_1$
	(0.000)	(0.000)	(0.000)	(**0.000**)
optdigi	0.035_2	$\mathbf{0.033}_1$	0.044_4	0.040_3
	(0.009)	(**0.006**)	(0.021)	(0.008)
page-bl	0.093_2	$\mathbf{0.089}_1$	0.098_3	0.109_4
	(0.014)	(**0.010**)	(0.028)	(0.013)
pendigi	0.017_2	$\mathbf{0.017}_1$	0.023_4	0.021_3
	(0.003)	(**0.003**)	(0.009)	(0.004)
scene-c	0.384_4	$\mathbf{0.364}_1$	0.376_2	0.378_3
	(0.030)	(**0.016**)	(0.038)	(0.017)
segment	0.020_3	$\mathbf{0.010}_1$	0.023_4	0.013_2
	(0.029)	(**0.009**)	(0.034)	(0.010)
shuttle	0.000_2	$\mathbf{0.000}_1$	0.001_4	0.000_3
	(0.001)	(**0.001**)	(0.001)	(0.001)
sonar	0.941_4	$\mathbf{0.404}_1$	0.486_3	0.440_2
	(0.539)	(**0.078**)	(0.332)	(0.071)
spambas	0.162_2	$\mathbf{0.162}_1$	0.173_4	0.167_3
	(0.019)	(**0.016**)	(0.035)	(0.016)
tic-tac	0.379_4	$\mathbf{0.333}_1$	0.340_3	0.339_2
	(0.021)	(**0.034**)	(0.060)	(0.035)
vehicle	0.077_2	$\mathbf{0.068}_1$	0.131_4	0.077_3
	(0.047)	(**0.021**)	(0.128)	(0.023)
vowel	0.076_2	$\mathbf{0.071}_1$	0.094_4	0.085_3
	(0.041)	(**0.021**)	(0.092)	(0.026)
wavefor	0.253_2	$\mathbf{0.252}_1$	0.264_3	0.271_4
	(0.018)	(**0.013**)	(0.025)	(0.016)
wdbc	0.256_4	$\mathbf{0.089}_1$	0.141_3	0.107_2
	(0.233)	(**0.036**)	(0.158)	(0.038)
wpbc	1.016_4	0.500_2	$\mathbf{0.496}_1$	0.503_3
	(0.366)	(0.039)	(**0.049**)	(0.029)
yeast	0.510_2	$\mathbf{0.509}_1$	0.543_4	0.514_3
	(0.028)	(**0.022**)	(0.089)	(0.022)
zoo	0.132_4	0.013_3	$\mathbf{0.013}_1$	0.013_2
	(0.011)	(0.020)	(**0.020**)	(0.020)
rank	3.20	1.27	3.12	2.41

Table 9: Brier score results for Ada-O (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $3.745e{-}06$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	0.213_2	$\mathbf{0.213}_1$	0.214_4	0.214_3
	(0.004)	(**0.004**)	(0.005)	(0.004)
autos	0.074_2	0.078_3	$\mathbf{0.071}_1$	0.080_4
	(0.030)	(0.024)	(**0.023**)	(0.025)
balance	0.013_4	0.010_2	0.011_3	$\mathbf{0.010}_1$
	(0.005)	(0.006)	(0.006)	(**0.006**)
car	0.034_3	0.034_2	$\mathbf{0.033}_1$	0.035_4
	(0.005)	(0.006)	(**0.006**)	(0.007)
clevela	0.133_3	0.131_2	$\mathbf{0.130}_1$	0.133_4
	(0.030)	(0.020)	(**0.020**)	(0.022)
credit-	$\mathbf{0.100}_1$	0.101_4	0.101_3	0.100_2
	(**0.018**)	(0.015)	(0.016)	(0.016)
dermato	0.003_4	0.003_2	$\mathbf{0.003}_1$	0.003_3
	(0.005)	(0.005)	(**0.005**)	(0.005)
diabete	0.164_4	$\mathbf{0.160}_1$	0.161_2	0.163_3
	(0.017)	(**0.011**)	(0.013)	(0.012)
ecoli	$\mathbf{0.036}_1$	0.037_2	0.037_3	0.037_4
	(**0.022**)	(0.020)	(0.019)	(0.020)
flare	0.130_2	$\mathbf{0.130}_1$	0.131_4	0.130_3
	(0.008)	(**0.008**)	(0.009)	(0.009)
german	0.166_2	0.167_3	0.167_4	$\mathbf{0.166}_1$
	(0.011)	(0.009)	(0.009)	(**0.008**)
glass	0.158_4	0.157_3	$\mathbf{0.153}_1$	0.154_2
	(0.037)	(0.023)	(**0.026**)	(0.023)
heart-s	0.141_4	0.134_2	$\mathbf{0.134}_1$	0.139_3
	(0.036)	(0.023)	(**0.026**)	(0.025)
hepatit	0.134_4	0.124_2	$\mathbf{0.123}_1$	0.128_3
	(0.047)	(0.022)	(**0.026**)	(0.023)
horse	0.130_4	$\mathbf{0.126}_1$	0.126_2	0.127_3
	(0.030)	(**0.021**)	(0.023)	(0.020)
ionosph	0.061_4	$\mathbf{0.057}_1$	0.058_2	0.059_3
	(0.018)	(**0.015**)	(0.016)	(0.016)
iris	0.014_4	0.000_2	$\mathbf{0.000}_1$	0.000_3
	(0.000)	(0.000)	(**0.000**)	(0.000)
landsat	0.012_3	$\mathbf{0.012}_1$	0.012_2	0.013_4
	(0.002)	(**0.002**)	(0.002)	(0.003)
letter	0.010_3	0.010_2	$\mathbf{0.010}_1$	0.011_4
	(0.001)	(0.001)	(**0.001**)	(0.001)
libras-	0.025_4	$\mathbf{0.022}_1$	0.022_2	0.024_3
	(0.014)	(**0.013**)	(0.013)	(0.013)
lung-ca	0.040_4	$\mathbf{0.032}_1$	0.032_3	0.032_2
	(0.020)	(**0.022**)	(0.022)	(0.022)
mfeat-k	0.006_3	$\mathbf{0.005}_1$	0.006_2	0.006_4
	(0.003)	(**0.003**)	(0.003)	(0.003)
mfeat-m	$\mathbf{0.002}_1$	0.002_2	0.002_4	0.002_3
	(**0.002**)	(0.002)	(0.002)	(0.002)
mfeat-z	0.007_3	$\mathbf{0.007}_1$	0.007_2	0.008_4
	(0.004)	(**0.003**)	(0.003)	(0.004)
mushroo	0.000_4	0.000_3	$\mathbf{0.000}_1$	0.000_2
	(0.000)	(0.000)	(**0.000**)	(0.000)
optdigi	0.008_2	$\mathbf{0.008}_1$	0.008_3	0.009_4
	(0.002)	(**0.002**)	(0.002)	(0.002)
page-bl	0.027_3	0.026_2	$\mathbf{0.026}_1$	0.028_4
	(0.004)	(0.003)	(**0.003**)	(0.004)
pendigi	0.005_2	$\mathbf{0.005}_1$	0.005_3	0.005_4
	(0.001)	(**0.001**)	(0.001)	(0.001)
scene-c	$\mathbf{0.115}_1$	0.115_3	0.115_2	0.116_4
	(**0.009**)	(0.006)	(0.006)	(0.007)
segment	0.002_3	$\mathbf{0.002}_1$	0.002_2	0.002_4
	(0.002)	(**0.001**)	(0.002)	(0.002)
shuttle	0.000_3	0.000_4	0.000_2	$\mathbf{0.000}_1$
	(0.000)	(0.000)	(0.000)	(**0.000**)
sonar	0.135_3	$\mathbf{0.127}_1$	0.130_2	0.136_4
	(0.043)	(**0.029**)	(0.031)	(0.030)
spambas	$\mathbf{0.044}_1$	0.045_4	0.045_2	0.045_3
	(**0.004**)	(0.004)	(0.004)	(0.005)
tic-tac	0.118_4	$\mathbf{0.107}_1$	0.108_2	0.108_3
	(0.010)	(**0.013**)	(0.013)	(0.014)
vehicle	$\mathbf{0.017}_1$	0.017_2	0.018_4	0.018_3
	(**0.006**)	(0.005)	(0.006)	(0.006)
vowel	0.020_3	0.019_2	$\mathbf{0.019}_1$	0.020_4
	(0.006)	(0.006)	(**0.006**)	(0.007)
wavefor	0.079_2	$\mathbf{0.079}_1$	0.079_3	0.081_4
	(0.006)	(**0.005**)	(0.005)	(0.006)
wdbc	0.024_3	$\mathbf{0.023}_1$	0.024_2	0.024_4
	(0.012)	(**0.011**)	(0.011)	(0.012)
wpbc	0.173_4	0.162_2	$\mathbf{0.160}_1$	0.163_3
	(0.032)	(0.015)	(**0.016**)	(0.012)
yeast	$\mathbf{0.173}_1$	0.173_2	0.174_4	0.173_3
	(**0.008**)	(0.008)	(0.009)	(0.009)
zoo	0.016_4	0.002_2	$\mathbf{0.002}_1$	0.002_3
	(0.004)	(0.003)	(**0.003**)	(0.003)
rank	2.85	1.85	2.12	3.17

Table 10: Accuracy results for Ada-O in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results not significant, according to Friedman's test (p-value = 0.813).

dataset	uncalibrated	beta	isotonic	logistic
abalone	63.907_3	63.969_2	63.009_4	$\mathbf{64.089}_1$
	(1.218)	(0.642)	(1.899)	(**0.932**)
autos	90.750_4	91.329_2	$\mathbf{91.866}_1$	90.821_3
	(5.060)	(5.280)	(**5.014**)	(5.024)
balance	98.606_4	98.703_2	$\mathbf{98.704}_1$	98.671_3
	(1.122)	(1.109)	(**1.106**)	(1.119)
car	95.226_2	95.162_3	$\mathbf{95.295}_1$	94.994_4
	(0.990)	(0.952)	(**1.077**)	(1.069)
clevela	$\mathbf{82.417}_1$	82.274_2	82.046_4	82.176_3
	(**4.493**)	(4.385)	(4.124)	(4.379)
credit-	$\mathbf{86.617}_1$	86.311_4	86.342_3	86.494_2
	(**2.766**)	(2.457)	(2.544)	(2.753)
dermato	$\mathbf{99.638}_1$	99.611_3	99.611_3	99.611_3
	(**0.681**)	(0.693)	(0.693)	(0.693)
diabete	$\mathbf{75.573}_1$	75.234_3	75.416_2	75.196_4
	(**3.176**)	(3.209)	(2.986)	(3.116)
ecoli	$\mathbf{95.815}_1$	95.700_2	95.695_3	95.639_4
	(**2.766**)	(2.834)	(2.669)	(2.863)
flare	$\mathbf{82.296}_1$	82.289_2	82.217_3	82.079_4
	(**1.122**)	(1.140)	(1.157)	(1.251)
german	$\mathbf{76.210}_1$	75.640_4	75.820_3	75.880_2
	(**1.990**)	(1.990)	(1.950)	(1.837)
glass	$\mathbf{80.508}_1$	79.671_3	79.564_4	80.436_2
	(**4.813**)	(4.888)	(4.653)	(4.857)
heart-s	81.333_4	81.444_2	$\mathbf{81.741}_1$	81.370_3
	(5.222)	(4.544)	(**5.075**)	(5.108)
hepatit	$\mathbf{82.783}_1$	81.647_4	81.883_2	81.879_3
	(**6.004**)	(4.494)	(5.051)	(4.452)
horse	$\mathbf{83.845}_1$	83.374_4	83.575_3	83.677_2
	(**4.487**)	(4.039)	(4.068)	(4.161)
ionosph	92.482_4	$\mathbf{93.052}_1$	92.909_2	92.711_3
	(2.483)	(**2.392**)	(2.474)	(2.489)
iris	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
landsat	$\mathbf{98.344}_1$	98.340_2	98.308_4	98.337_3
	(**0.355**)	(0.353)	(0.370)	(0.356)
letter	98.653_3	98.644_4	$\mathbf{98.730}_1$	98.656_2
	(0.134)	(0.129)	(**0.133**)	(0.130)
libras-	97.187_4	$\mathbf{97.613}_1$	97.607_2	97.307_3
	(1.437)	(**1.484**)	(1.517)	(1.441)
lung-ca	$\mathbf{97.295}_1$	95.521_3	95.521_3	95.521_3
	(**3.820**)	(4.457)	(4.457)	(4.457)
mfeat-k	99.315_4	$\mathbf{99.350}_1$	99.340_2	99.325_3
	(0.416)	(**0.378**)	(0.387)	(0.405)
mfeat-m	99.800_2	99.800_2	99.785_4	99.800_2
	(0.208)	(0.208)	(0.202)	(0.208)
mfeat-z	$99.065_{3.5}$	99.145_2	$\mathbf{99.155}_1$	$99.065_{3.5}$
	(0.522)	(0.446)	(**0.463**)	(0.487)
mushroo	$99.999_{3.5}$	$100.000_{1.5}$	$100.000_{1.5}$	$99.999_{3.5}$
	(0.009)	(0.000)	(0.000)	(0.009)
optdigi	98.924_3	98.924_2	$\mathbf{98.927}_1$	98.917_4
	(0.250)	(0.252)	(**0.269**)	(0.249)
page-bl	96.516_2	96.468_3	$\mathbf{96.552}_1$	96.450_4
	(0.481)	(0.512)	(**0.515**)	(0.549)
pendigi	99.399_4	99.411_2	99.402_3	$\mathbf{99.414}_1$
	(0.110)	(0.122)	(0.114)	(**0.124**)
scene-c	$\mathbf{84.130}_1$	83.801_4	83.847_3	83.984_2
	(**1.554**)	(1.480)	(1.548)	(1.462)
segment	99.758_2	99.740_4	99.749_3	$\mathbf{99.762}_1$
	(0.221)	(0.231)	(0.229)	(**0.220**)
shuttle	99.995_3	99.994_4	99.995_2	$\mathbf{99.995}_1$
	(0.006)	(0.006)	(0.006)	(**0.006**)
sonar	82.592_2	$\mathbf{82.727}_1$	81.709_4	82.493_3
	(6.334)	(**6.320**)	(6.047)	(6.438)
spambas	$\mathbf{94.371}_1$	94.358_2	94.343_4	94.353_3
	(**0.712**)	(0.736)	(0.698)	(0.735)
tic-tac	$\mathbf{83.913}_1$	83.831_3	83.778_4	83.841_2
	(**2.130**)	(2.678)	(2.722)	(2.680)
vehicle	97.814_3	97.849_2	97.743_4	$\mathbf{97.862}_1$
	(0.906)	(0.872)	(0.927)	(**0.934**)
vowel	97.404_4	$\mathbf{97.626}_1$	97.606_2	97.424_3
	(0.860)	(**0.891**)	(0.906)	(0.861)
wavefor	88.676_4	88.706_2	88.680_3	$\mathbf{88.722}_1$
	(0.989)	(0.938)	(0.895)	(**0.962**)
wdbc	96.889_4	96.923_3	96.924_2	$\mathbf{96.977}_1$
	(1.775)	(1.668)	(1.702)	(**1.748**)
wpbc	77.678_2	77.382_3	$\mathbf{78.615}_1$	77.020_4
	(4.915)	(3.023)	(**3.685**)	(2.498)
yeast	73.075_2	73.042_4	73.075_3	$\mathbf{73.137}_1$
	(2.200)	(2.137)	(2.219)	(**2.216**)
zoo	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
rank	2.34	2.55	2.52	2.59

Table 11: Log-loss results for Ada-O (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.598e{-}15$, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	0.612_4	0.612_2	$\mathbf{0.612}_1$	0.612_3	0.614_5
	(0.011)	(0.010)	(**0.010**)	(0.009)	(0.009)
autos	0.427_5	$\mathbf{0.283}_1$	0.288_3	0.292_4	0.287_2
	(0.359)	(**0.073**)	(0.072)	(0.070)	(0.066)
balance	0.049_5	$\mathbf{0.037}_1$	0.038_4	0.037_2	0.038_3
	(0.013)	(**0.018**)	(0.018)	(0.017)	(0.019)
car	0.114_4	0.109_2	$\mathbf{0.109}_1$	0.110_3	0.119_5
	(0.014)	(0.018)	(**0.017**)	(0.018)	(0.021)
clevela	0.496_5	0.417_3	0.416_2	$\mathbf{0.415}_1$	0.429_4
	(0.150)	(0.051)	(0.050)	(**0.049**)	(0.052)
credit-	0.360_5	0.341_4	0.339_3	0.338_2	$\mathbf{0.338}_1$
	(0.080)	(0.045)	(0.044)	(0.042)	(**0.041**)
dermato	0.036_5	0.019_2	0.020_3	$\mathbf{0.019}_1$	0.021_4
	(0.108)	(0.035)	(0.035)	(**0.035**)	(0.038)
diabete	0.506_5	0.484_2	$\mathbf{0.483}_1$	0.484_3	0.495_4
	(0.050)	(0.028)	(**0.028**)	(0.027)	(0.028)
ecoli	0.326_5	0.145_3	$\mathbf{0.142}_1$	0.142_2	0.159_4
	(0.372)	(0.072)	(**0.068**)	(0.068)	(0.073)
flare	$\mathbf{0.404}_1$	0.404_2	0.404_4	0.404_3	0.408_5
	(**0.023**)	(0.023)	(0.022)	(0.022)	(0.024)
german	0.511_5	0.506_3	0.506_4	$\mathbf{0.505}_1$	0.506_2
	(0.038)	(0.023)	(0.025)	(**0.023**)	(0.021)
glass	0.696_5	0.489_3	0.489_2	0.490_4	$\mathbf{0.485}_1$
	(0.266)	(0.059)	(0.059)	(0.057)	(**0.054**)
heart-s	0.570_5	0.427_3	0.426_2	$\mathbf{0.426}_1$	0.443_4
	(0.212)	(0.062)	(0.062)	(**0.061**)	(0.060)
hepatit	0.805_5	0.392_3	0.391_2	$\mathbf{0.390}_1$	0.411_4
	(0.507)	(0.056)	(0.055)	(**0.057**)	(0.061)
horse	0.642_5	$\mathbf{0.413}_1$	0.419_4	0.417_2	0.418_3
	(0.232)	(**0.053**)	(0.052)	(0.052)	(0.047)
ionosph	0.381_5	0.203_2	0.204_3	$\mathbf{0.201}_1$	0.227_4
	(0.203)	(0.044)	(0.043)	(**0.044**)	(0.045)
iris	0.127_5	$\mathbf{0.000}_1$	0.000_3	0.000_4	0.000_2
	(0.000)	(**0.000**)	(0.000)	(0.000)	(0.000)
landsat	0.041_4	0.040_2	0.040_3	$\mathbf{0.040}_1$	0.053_5
	(0.008)	(0.006)	(0.006)	(**0.006**)	(0.009)
letter	0.037_4	$\mathbf{0.035}_1$	0.036_2	0.036_3	0.044_5
	(0.002)	(**0.002**)	(0.002)	(0.002)	(0.003)
libras-	0.589_5	0.100_2	0.106_3	$\mathbf{0.098}_1$	0.112_4
	(0.502)	(0.054)	(0.052)	(**0.055**)	(0.053)
lung-ca	0.193_5	0.139_2	$\mathbf{0.138}_1$	0.139_3	0.139_4
	(0.051)	(0.178)	(**0.180**)	(0.180)	(0.184)
mfeat-k	0.062_5	0.027_2	$\mathbf{0.026}_1$	0.027_3	0.032_4
	(0.057)	(0.013)	(**0.012**)	(0.013)	(0.013)
mfeat-m	0.040_5	0.014_3	$\mathbf{0.012}_1$	0.014_2	0.014_4
	(0.060)	(0.017)	(**0.016**)	(0.016)	(0.016)
mfeat-z	0.095_5	0.032_3	0.032_2	$\mathbf{0.031}_1$	0.039_4
	(0.075)	(0.014)	(0.014)	(**0.014**)	(0.017)
mushroo	0.000_5	0.000_2	0.000_4	0.000_3	$\mathbf{0.000}_1$
	(0.000)	(0.000)	(0.000)	(0.000)	(**0.000**)
optdigi	0.035_4	0.033_2	$\mathbf{0.033}_1$	0.033_3	0.040_5
	(0.009)	(0.006)	(**0.006**)	(0.006)	(0.008)
page-bl	0.093_4	$\mathbf{0.089}_1$	0.090_3	0.090_2	0.109_5
	(0.014)	(**0.010**)	(0.010)	(0.010)	(0.013)
pendigi	0.017_2	$\mathbf{0.017}_1$	0.017_3	0.017_4	0.021_5
	(0.003)	(**0.003**)	(0.003)	(0.003)	(0.004)
scene-c	0.384_5	$\mathbf{0.364}_1$	0.365_3	0.365_2	0.378_4
	(0.030)	(**0.016**)	(0.016)	(0.016)	(0.017)
segment	0.020_5	0.010_3	0.010_2	$\mathbf{0.010}_1$	0.013_4
	(0.029)	(0.009)	(0.009)	(**0.009**)	(0.010)
shuttle	0.000_4	$\mathbf{0.000}_1$	0.000_3	0.000_2	0.000_5
	(0.001)	(**0.001**)	(0.001)	(0.001)	(0.001)
sonar	0.941_5	0.404_3	0.402_2	$\mathbf{0.401}_1$	0.440_4
	(0.539)	(0.078)	(0.075)	(**0.069**)	(0.071)
spambas	0.162_2	$\mathbf{0.162}_1$	0.162_3	0.163_4	0.167_5
	(0.019)	(**0.016**)	(0.016)	(0.016)	(0.016)
tic-tac	0.379_5	0.333_2	0.345_4	$\mathbf{0.333}_1$	0.339_3
	(0.021)	(0.034)	(0.032)	(**0.034**)	(0.035)
vehicle	0.077_4	0.068_3	0.067_2	$\mathbf{0.066}_1$	0.077_5
	(0.047)	(0.021)	(0.021)	(**0.020**)	(0.023)
vowel	0.076_4	$\mathbf{0.071}_1$	0.073_3	0.072_2	0.085_5
	(0.041)	(**0.021**)	(0.020)	(0.020)	(0.026)
wavefor	0.253_4	$\mathbf{0.252}_1$	0.252_2	0.252_3	0.271_5
	(0.018)	(**0.013**)	(0.013)	(0.013)	(0.016)
wdbc	0.256_5	0.089_2	$\mathbf{0.088}_1$	0.089_3	0.107_4
	(0.233)	(0.036)	(**0.036**)	(0.036)	(0.038)
wpbc	1.016_5	0.500_2	0.505_4	$\mathbf{0.500}_1$	0.503_3
	(0.366)	(0.039)	(0.042)	(**0.032**)	(0.029)
yeast	0.510_3	0.509_2	$\mathbf{0.509}_1$	0.510_4	0.514_5
	(0.028)	(0.022)	(**0.022**)	(0.022)	(0.022)
zoo	0.132_5	0.013_2	0.013_4	0.013_3	$\mathbf{0.013}_1$
	(0.011)	(0.020)	(0.020)	(0.020)	(**0.020**)
rank	4.46	2.02	2.46	2.24	3.80

Table 12: Brier score results for Ada-O (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.400e{-}06$, according to Friedman's test.

dataset	uncalibrated	beta	beta05	beta2	logistic
abalone	0.213_4	0.213_2	$\mathbf{0.213}_1$	0.213_3	0.214_5
	(0.004)	(0.004)	(**0.004**)	(0.004)	(0.004)
autos	$\mathbf{0.074}_1$	0.078_2	0.078_3	0.080_4	0.080_5
	(**0.030**)	(0.024)	(0.023)	(0.021)	(0.025)
balance	0.013_5	0.010_3	0.010_4	0.010_2	$\mathbf{0.010}_1$
	(0.005)	(0.006)	(0.006)	(0.006)	(**0.006**)
car	0.034_4	$\mathbf{0.034}_1$	0.034_2	0.034_3	0.035_5
	(0.005)	(**0.006**)	(0.006)	(0.006)	(0.007)
clevela	0.133_4	0.131_3	$\mathbf{0.130}_1$	0.130_2	0.133_5
	(0.030)	(0.020)	(**0.019**)	(0.019)	(0.022)
credit-	$\mathbf{0.100}_1$	0.101_5	0.101_4	0.101_3	0.100_2
	(**0.018**)	(0.015)	(0.014)	(0.014)	(0.016)
dermato	0.003_5	0.003_2	0.003_4	$\mathbf{0.003}_1$	0.003_3
	(0.005)	(0.005)	(0.005)	(**0.005**)	(0.005)
diabete	0.164_5	0.160_3	$\mathbf{0.160}_1$	0.160_2	0.163_4
	(0.017)	(0.011)	(**0.011**)	(0.011)	(0.012)
ecoli	0.036_3	0.037_4	$\mathbf{0.036}_1$	0.036_2	0.037_5
	(0.022)	(0.020)	(**0.019**)	(0.019)	(0.020)
flare	0.130_2	$\mathbf{0.130}_1$	0.130_4	0.130_3	0.130_5
	(0.008)	(**0.008**)	(0.008)	(0.008)	(0.009)
german	0.166_2	0.167_4	0.167_5	0.167_3	$\mathbf{0.166}_1$
	(0.011)	(0.009)	(0.009)	(0.008)	(**0.008**)
glass	0.158_5	0.157_3	0.157_2	0.157_4	$\mathbf{0.154}_1$
	(0.037)	(0.023)	(0.022)	(0.022)	(**0.023**)
heart-s	0.141_5	0.134_3	0.134_2	$\mathbf{0.134}_1$	0.139_4
	(0.036)	(0.023)	(0.023)	(**0.023**)	(0.025)
hepatit	0.134_5	0.124_3	0.124_2	$\mathbf{0.123}_1$	0.128_4
	(0.047)	(0.022)	(0.023)	(**0.023**)	(0.023)
horse	0.130_5	$\mathbf{0.126}_1$	0.128_4	0.127_2	0.127_3
	(0.030)	(**0.021**)	(0.020)	(0.019)	(0.020)
ionosph	0.061_5	0.057_2	0.057_3	$\mathbf{0.056}_1$	0.059_4
	(0.018)	(0.015)	(0.015)	(**0.015**)	(0.016)
iris	0.014_5	$\mathbf{0.000}_1$	0.000_3	0.000_4	0.000_2
	(0.000)	(**0.000**)	(0.000)	(0.000)	(0.000)
landsat	0.012_4	0.012_2	0.012_3	$\mathbf{0.012}_1$	0.013_5
	(0.002)	(0.002)	(0.002)	(**0.002**)	(0.003)
letter	0.010_4	0.010_2	$\mathbf{0.010}_1$	0.010_3	0.011_5
	(0.001)	(0.001)	(**0.001**)	(0.001)	(0.001)
libras-	0.025_5	0.022_2	0.023_3	$\mathbf{0.021}_1$	0.024_4
	(0.014)	(0.013)	(0.013)	(**0.013**)	(0.013)
lung-ca	0.040_5	0.032_3	$\mathbf{0.032}_1$	0.032_2	0.032_4
	(0.020)	(0.022)	(**0.022**)	(0.022)	(0.022)
mfeat-k	0.006_4	$\mathbf{0.005}_1$	0.005_2	0.006_3	0.006_5
	(0.003)	(**0.003**)	(0.003)	(0.002)	(0.003)
mfeat-m	0.002_2	0.002_4	$\mathbf{0.002}_1$	0.002_3	0.002_5
	(0.002)	(0.002)	(**0.002**)	(0.002)	(0.002)
mfeat-z	0.007_4	0.007_2	0.007_3	$\mathbf{0.007}_1$	0.008_5
	(0.004)	(0.003)	(0.003)	(**0.003**)	(0.004)
mushroo	0.000_5	0.000_2	0.000_4	0.000_3	$\mathbf{0.000}_1$
	(0.000)	(0.000)	(0.000)	(0.000)	(**0.000**)
optdigi	0.008_3	0.008_2	$\mathbf{0.008}_1$	0.008_4	0.009_5
	(0.002)	(0.002)	(**0.002**)	(0.002)	(0.002)
page-bl	0.027_4	0.026_2	0.027_3	$\mathbf{0.026}_1$	0.028_5
	(0.004)	(0.003)	(0.003)	(**0.003**)	(0.004)
pendigi	0.005_2	$\mathbf{0.005}_1$	0.005_4	0.005_3	0.005_5
	(0.001)	(**0.001**)	(0.001)	(0.001)	(0.001)
scene-c	$\mathbf{0.115}_1$	0.115_2	0.115_3	0.115_4	0.116_5
	(**0.009**)	(0.006)	(0.006)	(0.006)	(0.007)
segment	0.002_4	0.002_3	$\mathbf{0.002}_1$	0.002_2	0.002_5
	(0.002)	(0.001)	(**0.002**)	(0.001)	(0.002)
shuttle	0.000_2	0.000_3	0.000_4	0.000_5	$\mathbf{0.000}_1$
	(0.000)	(0.000)	(0.000)	(0.000)	(**0.000**)
sonar	0.135_4	0.127_3	$\mathbf{0.126}_1$	0.126_2	0.136_5
	(0.043)	(0.029)	(**0.028**)	(0.028)	(0.030)
spambas	$\mathbf{0.044}_1$	0.045_3	0.045_5	0.045_4	0.045_2
	(**0.004**)	(0.004)	(0.004)	(0.004)	(0.005)
tic-tac	0.118_5	$\mathbf{0.107}_1$	0.111_4	0.108_2	0.108_3
	(0.010)	(**0.013**)	(0.012)	(0.013)	(0.014)
vehicle	$\mathbf{0.017}_1$	0.017_4	0.017_3	0.017_2	0.018_5
	(**0.006**)	(0.005)	(0.006)	(0.005)	(0.006)
vowel	0.020_3	$\mathbf{0.019}_1$	0.020_4	0.020_2	0.020_5
	(0.006)	(**0.006**)	(0.006)	(0.006)	(0.007)
wavefor	0.079_4	$\mathbf{0.079}_1$	0.079_3	0.079_2	0.081_5
	(0.006)	(**0.005**)	(0.005)	(0.005)	(0.006)
wdbc	0.024_4	0.023_3	$\mathbf{0.023}_1$	0.023_2	0.024_5
	(0.012)	(0.011)	(**0.010**)	(0.011)	(0.012)
wpbc	0.173_5	$\mathbf{0.162}_1$	0.164_4	0.162_2	0.163_3
	(0.032)	(**0.015**)	(0.015)	(0.011)	(0.012)
yeast	$\mathbf{0.173}_1$	0.173_3	0.173_2	0.173_5	0.173_4
	(**0.008**)	(0.008)	(0.008)	(0.007)	(0.009)
zoo	0.016_5	0.002_2	$\mathbf{0.002}_1$	0.002_3	0.002_4
	(0.004)	(0.003)	(**0.003**)	(0.003)	(0.003)
rank	3.61	2.34	2.63	2.51	3.90

Table 13: Accuracy results for Ada-O in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at 0.048, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	63.907_4	63.969_3	63.823_5	$\mathbf{64.232}_1$	64.089_2
	(1.218)	(0.642)	(1.050)	(**0.825**)	(0.932)
autos	90.750_4	$\mathbf{91.329}_1$	91.174_2	90.676_5	90.821_3
	(5.060)	(**5.280**)	(4.829)	(4.626)	(5.024)
balance	98.606_5	$\mathbf{98.703}_1$	98.672_3	98.703_2	98.671_4
	(1.122)	(**1.109**)	(1.017)	(1.132)	(1.119)
car	$\mathbf{95.226}_1$	95.162_4	95.191_2	95.168_3	94.994_5
	(**0.990**)	(0.952)	(0.969)	(1.063)	(1.069)
clevela	82.417_2	82.274_4	$\mathbf{82.580}_1$	82.279_3	82.176_5
	(4.493)	(4.385)	(**4.578**)	(4.219)	(4.379)
credit-	$\mathbf{86.617}_1$	86.311_5	86.570_2	86.463_4	86.494_3
	(**2.766**)	(2.457)	(2.570)	(2.505)	(2.753)
dermato	$99.638_{1.5}$	99.611_4	$99.638_{1.5}$	99.611_4	99.611_4
	(0.681)	(0.693)	(0.681)	(0.693)	(0.693)
diabete	$\mathbf{75.573}_1$	75.234_4	75.404_2	75.326_3	75.196_5
	(**3.176**)	(3.209)	(3.142)	(3.110)	(3.116)
ecoli	95.815_2	95.700_4	$\mathbf{95.816}_1$	95.788_3	95.639_5
	(2.766)	(2.834)	(**2.658**)	(2.757)	(2.863)
flare	82.296_2	82.289_4	82.296_2	82.296_2	82.079_5
	(1.122)	(1.140)	(1.122)	(1.122)	(1.251)
german	76.210_2	75.640_4	$\mathbf{76.280}_1$	75.580_5	75.880_3
	(1.990)	(1.990)	(**1.721**)	(2.014)	(1.837)
glass	$\mathbf{80.508}_1$	79.671_5	79.930_4	80.223_3	80.436_2
	(**4.813**)	(4.888)	(4.017)	(4.607)	(4.857)
heart-s	81.333_5	81.444_3	81.667_2	$\mathbf{82.185}_1$	81.370_4
	(5.222)	(4.544)	(5.023)	(**4.641**)	(5.108)
hepatit	$\mathbf{82.783}_1$	81.647_5	82.341_2	82.089_3	81.879_4
	(**6.004**)	(4.494)	(5.233)	(4.666)	(4.452)
horse	$\mathbf{83.845}_1$	83.374_5	83.648_3	83.607_4	83.677_2
	(**4.487**)	(4.039)	(4.295)	(3.816)	(4.161)
ionosph	92.482_5	93.052_2	92.737_3	$\mathbf{93.080}_1$	92.711_4
	(2.483)	(2.392)	(2.540)	(**2.546**)	(2.489)
iris	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
landsat	98.344_2	98.340_3	$\mathbf{98.368}_1$	98.339_4	98.337_5
	(0.355)	(0.353)	(**0.361**)	(0.361)	(0.356)
letter	98.653_3	98.644_4	98.654_2	98.599_5	$\mathbf{98.656}_1$
	(0.134)	(0.129)	(0.128)	(0.132)	(**0.130**)
libras-	97.187_5	97.613_2	97.333_3	$\mathbf{97.720}_1$	97.307_4
	(1.437)	(1.484)	(1.454)	(**1.532**)	(1.441)
lung-ca	$\mathbf{97.295}_1$	$95.521_{4.5}$	$95.621_{2.5}$	$95.621_{2.5}$	$95.521_{4.5}$
	(**3.820**)	(4.457)	(4.386)	(4.386)	(4.457)
mfeat-k	99.315_5	99.350_2	99.335_3	$\mathbf{99.350}_1$	99.325_4
	(0.416)	(0.378)	(0.396)	(**0.354**)	(0.405)
mfeat-m	$99.800_{2.5}$	$99.800_{2.5}$	$99.800_{2.5}$	99.795_5	$99.800_{2.5}$
	(0.208)	(0.208)	(0.208)	(0.206)	(0.208)
mfeat-z	$99.065_{4.5}$	$\mathbf{99.145}_1$	99.095_3	99.135_2	$99.065_{4.5}$
	(0.522)	(**0.446**)	(0.507)	(0.463)	(0.487)
mushroo	99.999_4	$100.000_{1.5}$	99.999_4	$100.000_{1.5}$	99.999_4
	(0.009)	(0.000)	(0.009)	(0.000)	(0.009)
optdigi	98.924_3	98.924_2	$\mathbf{98.925}_1$	98.924_4	98.917_5
	(0.250)	(0.252)	(**0.239**)	(0.238)	(0.249)
page-bl	$\mathbf{96.516}_1$	96.468_4	96.512_2	96.503_3	96.450_5
	(**0.481**)	(0.512)	(0.494)	(0.499)	(0.549)
pendigi	99.399_5	99.411_2	99.402_4	99.403_3	$\mathbf{99.414}_1$
	(0.110)	(0.122)	(0.116)	(0.125)	(**0.124**)
scene-c	$\mathbf{84.130}_1$	83.801_5	83.972_3	83.855_4	83.984_2
	(**1.554**)	(1.480)	(1.599)	(1.446)	(1.462)
segment	99.758_3	99.740_5	$\mathbf{99.771}_1$	99.753_4	99.762_2
	(0.221)	(0.231)	(**0.198**)	(0.223)	(0.220)
shuttle	99.995_3	99.994_4	99.995_2	99.994_5	$\mathbf{99.995}_1$
	(0.006)	(0.006)	(0.006)	(0.006)	(**0.006**)
sonar	82.592_4	82.727_2	82.724_3	$\mathbf{82.965}_1$	82.493_5
	(6.334)	(6.320)	(6.563)	(**6.682**)	(6.438)
spambas	94.371_2	94.358_4	94.371_3	$\mathbf{94.375}_1$	94.353_5
	(0.712)	(0.736)	(0.713)	(**0.719**)	(0.735)
tic-tac	83.913_2	83.831_4	$\mathbf{84.164}_1$	83.810_5	83.841_3
	(2.130)	(2.678)	(**2.088**)	(2.731)	(2.680)
vehicle	97.814_5	97.849_4	97.874_2	$\mathbf{97.956}_1$	97.862_3
	(0.906)	(0.872)	(0.941)	(**0.850**)	(0.934)
vowel	$97.404_{4.5}$	$\mathbf{97.626}_1$	$97.404_{4.5}$	97.525_2	97.424_3
	(0.860)	(**0.891**)	(0.890)	(0.959)	(0.861)
wavefor	88.676_3	88.706_2	88.670_5	88.674_4	$\mathbf{88.722}_1$
	(0.989)	(0.938)	(0.970)	(0.969)	(**0.962**)
wdbc	96.889_5	96.923_4	96.976_2	96.924_3	$\mathbf{96.977}_1$
	(1.775)	(1.668)	(1.659)	(1.722)	(**1.748**)
wpbc	77.678_2	77.382_3	$\mathbf{77.937}_1$	76.510_5	77.020_4
	(4.915)	(3.023)	(**3.707**)	(1.889)	(2.498)
yeast	73.075_3	73.042_4	73.082_2	72.704_5	$\mathbf{73.137}_1$
	(2.200)	(2.137)	(2.233)	(2.146)	(**2.216**)
zoo	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
rank	2.88	3.28	2.44	3.05	3.35

Table 14: Log-loss results for Ada-S (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.008e{-}12$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	0.691_4	0.632_2	$\mathbf{0.620}_1$	0.635_3
	(0.000)	(0.007)	(**0.022**)	(0.008)
autos	0.517_4	$\mathbf{0.278}_1$	0.498_3	0.282_2
	(0.022)	(**0.082**)	(0.580)	(0.068)
balance	0.643_4	$\mathbf{0.032}_1$	0.041_3	0.032_2
	(0.003)	(**0.023**)	(0.085)	(0.023)
car	0.650_4	$\mathbf{0.108}_1$	0.141_3	0.108_2
	(0.002)	(**0.017**)	(0.071)	(0.017)
clevela	0.670_4	$\mathbf{0.482}_1$	0.516_3	0.484_2
	(0.003)	(**0.046**)	(0.221)	(0.040)
credit-	0.675_4	$\mathbf{0.403}_1$	0.426_3	0.412_2
	(0.002)	(**0.039**)	(0.119)	(0.044)
dermato	0.044_2	$\mathbf{0.042}_1$	0.112_4	0.051_3
	(0.101)	(**0.077**)	(0.292)	(0.096)
diabete	0.683_4	$\mathbf{0.537}_1$	0.556_3	0.554_2
	(0.002)	(**0.020**)	(0.115)	(0.023)
ecoli	0.427_4	0.149_2	0.279_3	$\mathbf{0.146}_1$
	(0.053)	(0.074)	(0.361)	(**0.069**)
flare	0.646_4	$\mathbf{0.415}_1$	0.426_3	0.416_2
	(0.013)	(**0.021**)	(0.077)	(0.021)
german	0.688_4	0.550_2	$\mathbf{0.542}_1$	0.563_3
	(0.001)	(0.018)	(**0.042**)	(0.016)
glass	0.659_4	$\mathbf{0.502}_1$	0.524_3	0.516_2
	(0.006)	(**0.044**)	(0.186)	(0.044)
heart-s	0.665_4	$\mathbf{0.479}_1$	0.539_3	0.484_2
	(0.004)	(**0.051**)	(0.223)	(0.046)
hepatit	0.562_4	0.389_2	0.424_3	$\mathbf{0.389}_1$
	(0.025)	(0.053)	(0.236)	(**0.051**)
horse	0.644_4	$\mathbf{0.434}_1$	0.510_3	0.444_2
	(0.007)	(**0.047**)	(0.230)	(0.046)
ionosph	0.531_4	$\mathbf{0.199}_1$	0.267_3	0.199_2
	(0.015)	(**0.048**)	(0.237)	(0.046)
iris	0.000_2	0.000_3	$\mathbf{0.000}_1$	0.000_4
	(0.000)	(0.000)	(**0.000**)	(0.000)
landsat	0.608_4	$\mathbf{0.038}_1$	0.040_3	0.038_2
	(0.006)	(**0.008**)	(0.019)	(0.008)
letter	0.577_4	0.022_2	$\mathbf{0.021}_1$	0.022_3
	(0.003)	(0.003)	(**0.003**)	(0.003)
libras-	0.237_4	$\mathbf{0.098}_1$	0.168_3	0.099_2
	(0.036)	(**0.064**)	(0.254)	(0.053)
lung-ca	0.218_4	$\mathbf{0.138}_1$	0.189_3	0.140_2
	(0.632)	(**0.178**)	(0.534)	(0.187)
mfeat-k	0.460_4	$\mathbf{0.029}_1$	0.044_3	0.030_2
	(0.020)	(**0.013**)	(0.052)	(0.013)
mfeat-m	0.076_4	$\mathbf{0.014}_1$	0.029_3	0.015_2
	(0.029)	(**0.017**)	(0.043)	(0.016)
mfeat-z	0.351_4	$\mathbf{0.032}_1$	0.059_3	0.033_2
	(0.029)	(**0.014**)	(0.042)	(0.014)
mushroo	0.541_4	0.000_3	$\mathbf{0.000}_1$	0.000_2
	(0.005)	(0.000)	(**0.000**)	(0.000)
optdigi	0.577_4	$\mathbf{0.034}_1$	0.041_3	0.036_2
	(0.010)	(**0.007**)	(0.020)	(0.008)
page-bl	0.662_4	0.088_2	$\mathbf{0.086}_1$	0.093_3
	(0.001)	(0.011)	(**0.017**)	(0.013)
pendigi	0.474_4	$\mathbf{0.014}_1$	0.017_3	0.015_2
	(0.010)	(**0.004**)	(0.008)	(0.004)
scene-c	0.654_4	0.386_2	$\mathbf{0.381}_1$	0.394_3
	(0.006)	(0.012)	(**0.016**)	(0.011)
segment	0.241_4	0.009_2	0.023_3	$\mathbf{0.009}_1$
	(0.025)	(0.008)	(0.040)	(**0.007**)
shuttle	0.478_4	$\mathbf{0.000}_1$	0.000_3	0.000_2
	(0.020)	(**0.001**)	(0.001)	(0.001)
sonar	0.620_4	0.414_2	0.538_3	$\mathbf{0.413}_1$
	(0.009)	(0.073)	(0.389)	(**0.071**)
spambas	0.646_4	0.178_2	$\mathbf{0.159}_1$	0.182_3
	(0.004)	(0.013)	(**0.017**)	(0.014)
tic-tac	0.689_4	$\mathbf{0.169}_1$	0.181_3	0.170_2
	(0.000)	(**0.031**)	(0.097)	(0.030)
vehicle	0.575_4	$\mathbf{0.070}_1$	0.137_3	0.071_2
	(0.009)	(**0.021**)	(0.140)	(0.020)
vowel	0.422_4	$\mathbf{0.064}_1$	0.082_3	0.065_2
	(0.011)	(**0.019**)	(0.061)	(0.018)
wavefor	0.669_4	0.262_2	$\mathbf{0.259}_1$	0.266_3
	(0.004)	(0.011)	(**0.013**)	(0.012)
wdbc	0.422_4	$\mathbf{0.090}_1$	0.148_3	0.090_2
	(0.027)	(**0.033**)	(0.147)	(0.032)
wpbc	0.625_4	0.503_3	$\mathbf{0.501}_1$	0.502_2
	(0.012)	(0.036)	(**0.048**)	(0.033)
yeast	0.673_4	0.540_2	$\mathbf{0.527}_1$	0.541_3
	(0.002)	(0.012)	(**0.040**)	(0.011)
zoo	$\mathbf{0.006}_1$	0.013_4	0.013_2	0.013_3
	(**0.013**)	(0.020)	(0.020)	(0.020)
rank	3.83	1.51	2.46	2.20

Table 15: Brier score results for Ada-S (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $5.836e{-}13$, according to Friedman's test.

dataset	uncalibrated	beta	isotonic	logistic
abalone	0.249_4	0.220_2	$\mathbf{0.214}_1$	0.222_3
	(0.000)	(0.003)	(**0.003**)	(0.003)
autos	0.165_4	0.074_2	$\mathbf{0.069}_1$	0.076_3
	(0.010)	(0.025)	(**0.024**)	(0.022)
balance	0.225_4	$\mathbf{0.009}_1$	0.009_3	0.009_2
	(0.001)	(**0.006**)	(0.006)	(0.006)
car	0.229_4	0.033_2	0.034_3	$\mathbf{0.033}_1$
	(0.001)	(0.006)	(0.006)	(**0.006**)
clevela	0.238_4	0.155_2	$\mathbf{0.151}_1$	0.156_3
	(0.002)	(0.018)	(**0.021**)	(0.017)
credit-	0.241_4	0.121_2	$\mathbf{0.115}_1$	0.124_3
	(0.001)	(0.013)	(**0.013**)	(0.013)
dermato	$\mathbf{0.006}_1$	0.006_3	0.007_4	0.006_2
	(**0.009**)	(0.009)	(0.009)	(0.009)
diabete	0.245_4	0.178_2	$\mathbf{0.172}_1$	0.184_3
	(0.001)	(0.007)	(**0.011**)	(0.007)
ecoli	0.125_4	0.037_2	0.037_3	$\mathbf{0.037}_1$
	(0.021)	(0.019)	(0.019)	(**0.019**)
flare	0.227_4	0.133_2	$\mathbf{0.131}_1$	0.133_3
	(0.006)	(0.008)	(**0.009**)	(0.008)
german	0.248_4	0.184_2	$\mathbf{0.178}_1$	0.190_3
	(0.000)	(0.007)	(**0.008**)	(0.006)
glass	0.233_4	0.162_2	$\mathbf{0.156}_1$	0.167_3
	(0.003)	(0.019)	(**0.023**)	(0.017)
heart-s	0.236_4	0.154_2	$\mathbf{0.154}_1$	0.155_3
	(0.002)	(0.020)	(**0.025**)	(0.019)
hepatit	0.186_4	0.124_3	$\mathbf{0.122}_1$	0.123_2
	(0.012)	(0.020)	(**0.026**)	(0.020)
horse	0.226_4	0.133_2	$\mathbf{0.132}_1$	0.136_3
	(0.003)	(0.019)	(**0.022**)	(0.018)
ionosph	0.171_4	$\mathbf{0.054}_1$	0.054_3	0.054_2
	(0.007)	(**0.015**)	(0.016)	(0.015)
iris	0.000_2	0.000_3	$\mathbf{0.000}_1$	0.000_4
	(0.000)	(0.000)	(**0.000**)	(0.000)
landsat	0.208_4	$\mathbf{0.010}_1$	0.010_2	0.010_3
	(0.003)	(**0.002**)	(0.002)	(0.002)
letter	0.193_4	0.006_2	$\mathbf{0.006}_1$	0.006_3
	(0.001)	(0.001)	(**0.001**)	(0.001)
libras-	0.052_4	0.021_2	$\mathbf{0.021}_1$	0.021_3
	(0.010)	(0.013)	(**0.013**)	(0.012)
lung-ca	$\mathbf{0.027}_1$	0.032_2	0.032_4	0.032_3
	(**0.024**)	(0.022)	(0.022)	(0.022)
mfeat-k	0.137_4	$\mathbf{0.006}_1$	0.006_3	0.006_2
	(0.009)	(**0.003**)	(0.003)	(0.003)
mfeat-m	0.009_4	$\mathbf{0.002}_1$	0.003_3	0.002_2
	(0.004)	(**0.002**)	(0.002)	(0.002)
mfeat-z	0.091_4	$\mathbf{0.007}_1$	0.007_3	0.007_2
	(0.011)	(**0.003**)	(0.003)	(0.003)
mushroo	0.175_4	0.000_2	$\mathbf{0.000}_1$	0.000_3
	(0.002)	(0.000)	(**0.000**)	(0.000)
optdigi	0.192_4	0.008_2	$\mathbf{0.008}_1$	0.008_3
	(0.005)	(0.002)	(**0.002**)	(0.002)
page-bl	0.235_4	0.024_2	$\mathbf{0.023}_1$	0.025_3
	(0.001)	(0.003)	(**0.003**)	(0.003)
pendigi	0.145_4	0.003_2	$\mathbf{0.003}_1$	0.003_3
	(0.005)	(0.001)	(**0.001**)	(0.001)
scene-c	0.231_4	0.123_2	$\mathbf{0.120}_1$	0.126_3
	(0.003)	(0.005)	(**0.005**)	(0.004)
segment	0.050_4	0.002_2	0.002_3	$\mathbf{0.002}_1$
	(0.008)	(0.001)	(0.001)	(**0.001**)
shuttle	0.146_4	$\mathbf{0.000}_1$	0.000_3	0.000_2
	(0.009)	(**0.000**)	(0.000)	(0.000)
sonar	0.214_4	$\mathbf{0.131}_1$	0.132_3	0.131_2
	(0.004)	(**0.029**)	(0.031)	(0.029)
spambas	0.227_4	0.047_2	$\mathbf{0.043}_1$	0.047_3
	(0.002)	(0.003)	(**0.004**)	(0.003)
tic-tac	0.248_4	0.045_2	$\mathbf{0.043}_1$	0.045_3
	(0.000)	(0.009)	(**0.010**)	(0.009)
vehicle	0.192_4	$\mathbf{0.018}_1$	0.018_3	0.018_2
	(0.004)	(**0.005**)	(0.006)	(0.005)
vowel	0.123_4	0.017_2	$\mathbf{0.017}_1$	0.017_3
	(0.004)	(0.006)	(**0.006**)	(0.005)
wavefor	0.238_4	0.080_2	$\mathbf{0.079}_1$	0.081_3
	(0.002)	(0.004)	(**0.004**)	(0.004)
wdbc	0.121_4	$\mathbf{0.024}_1$	0.024_2	0.024_3
	(0.011)	(**0.010**)	(0.010)	(0.009)
wpbc	0.216_4	0.164_3	$\mathbf{0.163}_1$	0.164_2
	(0.006)	(0.013)	(**0.017**)	(0.012)
yeast	0.240_4	0.185_2	$\mathbf{0.175}_1$	0.186_3
	(0.001)	(0.004)	(**0.006**)	(0.003)
zoo	$\mathbf{0.002}_1$	0.002_4	0.002_2	0.002_3
	(**0.003**)	(0.003)	(0.003)	(0.003)
rank	3.73	1.90	1.76	2.61

Table 16: Accuracy results for Ada-S in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results not significant, according to Friedman's test (p-value = 0.660).

dataset	uncalibrated	beta	isotonic	logistic
abalone	63.304_4	$\mathbf{63.778}_1$	63.636_3	63.696_2
	(1.430)	(**0.329**)	(0.803)	(0.308)
autos	91.495_4	91.609_3	$\mathbf{92.006}_1$	91.674_2
	(4.969)	(4.483)	(**4.291**)	(4.618)
balance	98.830_4	98.975_3	98.976_2	$\mathbf{99.023}_1$
	(1.015)	(1.097)	(1.058)	(**1.064**)
car	95.034_4	95.156_3	$\mathbf{95.168}_1$	95.156_2
	(1.059)	(1.101)	(**1.072**)	(1.111)
clevela	$\mathbf{79.483}_1$	78.227_4	78.532_2	78.363_3
	(**4.701**)	(5.020)	(5.020)	(5.046)
credit-	84.258_2	83.968_3	$\mathbf{84.382}_1$	83.877_4
	(2.464)	(2.696)	(**2.461**)	(2.735)
dermato	99.408_2	99.408_2	99.297_4	99.408_2
	(1.111)	(1.111)	(1.185)	(1.111)
diabete	$\mathbf{74.115}_1$	73.591_3	73.945_2	72.070_4
	(**2.873**)	(2.166)	(2.649)	(2.658)
ecoli	$\mathbf{96.083}_1$	95.725_4	95.791_2	95.783_3
	(**2.450**)	(2.574)	(2.485)	(2.523)
flare	82.145_4	$82.260_{1.5}$	82.195_3	$82.260_{1.5}$
	(1.267)	(1.135)	(1.264)	(1.135)
german	$\mathbf{73.810}_1$	71.530_3	73.470_2	70.560_4
	(**2.470**)	(2.088)	(1.968)	(1.155)
glass	$\mathbf{80.460}_1$	79.168_3	79.518_2	78.516_4
	(**4.919**)	(5.608)	(5.074)	(4.386)
heart-s	$\mathbf{79.074}_1$	78.222_3	77.963_4	78.333_2
	(**5.347**)	(5.083)	(5.603)	(5.360)
hepatit	$\mathbf{82.916}_1$	80.733_4	82.318_2	81.697_3
	(**6.088**)	(4.211)	(5.145)	(4.561)
horse	$\mathbf{83.870}_1$	83.038_4	83.065_3	83.373_2
	(**4.325**)	(3.929)	(4.199)	(3.849)
ionosph	$\mathbf{93.336}_1$	93.223_3	93.194_4	93.336_2
	(**2.652**)	(2.511)	(2.539)	(2.392)
iris	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
landsat	98.636_4	98.645_3	$\mathbf{98.662}_1$	98.651_2
	(0.343)	(0.336)	(**0.337**)	(0.347)
letter	$\mathbf{99.202}_1$	99.200_2	99.200_3	99.198_4
	(**0.099**)	(0.102)	(0.098)	(0.102)
libras-	97.080_4	$\mathbf{97.693}_1$	97.607_3	97.640_2
	(1.580)	(**1.809**)	(1.802)	(1.811)
lung-ca	$\mathbf{97.295}_1$	95.621_2	$95.521_{3.5}$	$95.521_{3.5}$
	(**3.820**)	(4.386)	(4.457)	(4.457)
mfeat-k	$\mathbf{99.330}_1$	99.325_2	99.280_4	99.315_3
	(**0.393**)	(0.347)	(0.345)	(0.349)
mfeat-m	99.765_4	99.775_2	$\mathbf{99.790}_1$	99.770_3
	(0.211)	(0.216)	(**0.210**)	(0.213)
mfeat-z	99.155_4	99.185_3	99.185_2	$\mathbf{99.195}_1$
	(0.413)	(0.413)	(0.403)	(**0.402**)
mushroo	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
optdigi	98.899_4	98.925_2	$\mathbf{98.941}_1$	98.918_3
	(0.289)	(0.281)	(**0.286**)	(0.289)
page-bl	96.996_2	96.938_3	$\mathbf{97.047}_1$	96.867_4
	(0.461)	(0.426)	(**0.489**)	(0.470)
pendigi	99.674_2	99.671_3	$\mathbf{99.690}_1$	99.660_4
	(0.115)	(0.118)	(**0.107**)	(0.122)
scene-c	$\mathbf{83.224}_1$	82.526_3	83.057_2	81.495_4
	(**1.450**)	(1.168)	(1.388)	(1.105)
segment	99.732_4	99.740_2	99.732_3	$\mathbf{99.753}_1$
	(0.250)	(0.219)	(0.217)	(**0.205**)
shuttle	99.995_4	99.998_2	99.997_3	$\mathbf{99.998}_1$
	(0.006)	(0.005)	(0.006)	(**0.005**)
sonar	81.865_3	81.893_2	$\mathbf{82.101}_1$	81.851_4
	(6.839)	(6.172)	(**6.085**)	(6.459)
spambas	94.256_4	94.523_2	$\mathbf{94.590}_1$	94.508_3
	(0.637)	(0.618)	(**0.571**)	(0.573)
tic-tac	91.618_4	$\mathbf{95.189}_1$	95.074_3	95.178_2
	(2.118)	(**1.469**)	(1.495)	(1.458)
vehicle	97.766_3	97.849_2	97.672_4	$\mathbf{97.861}_1$
	(0.941)	(1.009)	(0.970)	(**0.967**)
vowel	97.455_4	97.960_2	$\mathbf{98.030}_1$	97.909_3
	(0.866)	(1.030)	(**0.903**)	(0.973)
wavefor	88.830_4	88.986_2	88.974_3	$\mathbf{88.996}_1$
	(0.788)	(0.773)	(0.720)	(**0.827**)
wdbc	$\mathbf{97.082}_1$	97.011_2	96.801_4	96.958_3
	(**1.437**)	(1.668)	(1.833)	(1.643)
wpbc	$\mathbf{77.806}_1$	76.605_3	76.876_2	76.457_4
	(**4.605**)	(2.390)	(4.008)	(2.117)
yeast	$\mathbf{72.639}_1$	69.177_4	72.143_2	69.251_3
	(**2.076**)	(0.654)	(1.689)	(0.814)
zoo	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$	$100.000_{2.5}$
	(0.000)	(0.000)	(0.000)	(0.000)
rank	2.48	2.56	2.32	2.65

Table 17: Log-loss results for Ada-S (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $2.799e{-}17$, according to Friedman's test.

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	0.691_5	$\mathbf{0.632_1}$	0.632_2	0.635_3	0.635_4
	(0.000)	(**0.007**)	(0.011)	(0.008)	(0.008)
autos	0.517_5	0.278_2	$\mathbf{0.277_1}$	0.285_4	0.282_3
	(0.022)	(0.082)	(**0.077**)	(0.069)	(0.068)
balance	0.643_5	$\mathbf{0.032_1}$	0.036_4	0.034_3	0.032_2
	(0.003)	(**0.023**)	(0.030)	(0.033)	(0.023)
car	0.650_5	0.108_3	$\mathbf{0.108_1}$	0.108_2	0.108_4
	(0.002)	(0.017)	(**0.017**)	(0.017)	(0.017)
clevela	0.670_5	$\mathbf{0.482_1}$	0.483_2	0.484_4	0.484_3
	(0.003)	(**0.046**)	(0.045)	(0.040)	(0.040)
credit-	0.675_5	0.403_2	$\mathbf{0.403_1}$	0.412_4	0.412_3
	(0.002)	(0.039)	(**0.038**)	(0.044)	(0.044)
dermato	0.044_3	0.042_2	$\mathbf{0.041_1}$	0.045_4	0.051_5
	(0.101)	(0.077)	(**0.074**)	(0.082)	(0.096)
diabete	0.683_5	$\mathbf{0.537_1}$	0.543_2	0.554_4	0.554_3
	(0.002)	(**0.020**)	(0.023)	(0.023)	(0.023)
ecoli	0.427_5	0.149_4	0.147_3	0.147_2	$\mathbf{0.146_1}$
	(0.053)	(0.074)	(0.065)	(0.068)	(**0.069**)
flare	0.646_5	$\mathbf{0.415_1}$	0.433_4	0.415_2	0.416_3
	(0.013)	(**0.021**)	(0.033)	(0.021)	(0.021)
german	0.688_5	$\mathbf{0.550_1}$	0.557_2	0.563_3	0.563_4
	(0.001)	(**0.018**)	(0.023)	(0.016)	(0.016)
glass	0.659_5	$\mathbf{0.502_1}$	0.504_2	0.516_4	0.516_3
	(0.006)	(**0.044**)	(0.045)	(0.044)	(0.044)
heart-s	0.665_5	$\mathbf{0.479_1}$	0.480_2	0.484_4	0.484_3
	(0.004)	(**0.051**)	(0.048)	(0.046)	(0.046)
hepatit	0.562_5	0.389_3	0.392_4	$\mathbf{0.389_1}$	0.389_2
	(0.025)	(0.053)	(0.054)	(**0.051**)	(0.051)
horse	0.644_5	$\mathbf{0.434_1}$	0.446_4	0.444_3	0.444_2
	(0.007)	(**0.047**)	(0.046)	(0.046)	(0.046)
ionosph	0.531_5	0.199_2	$\mathbf{0.198_1}$	0.199_4	0.199_3
	(0.015)	(0.048)	(**0.046**)	(0.046)	(0.046)
iris	$\mathbf{0.000_1}$	$0.000_{2.5}$	$0.000_{2.5}$	0.000_5	0.000_4
	(**0.000**)	(0.000)	(0.000)	(0.000)	(0.000)
landsat	0.608_5	0.038_2	$\mathbf{0.037_1}$	0.038_3	0.038_4
	(0.006)	(0.008)	(**0.007**)	(0.008)	(0.008)
letter	0.577_5	0.022_2	$\mathbf{0.022_1}$	0.022_3	0.022_4
	(0.003)	(0.003)	(**0.003**)	(0.003)	(0.003)
libras-	0.237_5	$\mathbf{0.098_1}$	0.103_3	0.111_4	0.099_2
	(0.036)	(**0.064**)	(0.054)	(0.063)	(0.053)
lung-ca	0.218_5	0.138_3	0.138_2	$\mathbf{0.133_1}$	0.140_4
	(0.632)	(0.178)	(0.178)	(**0.146**)	(0.187)
mfeat-k	0.460_5	$\mathbf{0.029_1}$	0.030_2	0.031_4	0.030_3
	(0.020)	(**0.013**)	(0.013)	(0.013)	(0.013)
mfeat-m	0.076_5	0.014_3	$\mathbf{0.013_1}$	0.014_2	0.015_4
	(0.029)	(0.017)	(**0.015**)	(0.016)	(0.016)
mfeat-z	0.351_5	0.032_2	$\mathbf{0.032_1}$	0.033_4	0.033_3
	(0.029)	(0.014)	(**0.014**)	(0.014)	(0.014)
mushroo	0.541_5	0.000_3	0.000_2	0.000_4	$\mathbf{0.000_1}$
	(0.005)	(0.000)	(0.000)	(0.000)	(**0.000**)
optdigi	0.577_5	0.034_2	$\mathbf{0.034_1}$	0.036_4	0.036_3
	(0.010)	(0.007)	(**0.007**)	(0.008)	(0.008)
page-bl	0.662_5	0.088_2	$\mathbf{0.088_1}$	0.093_3	0.093_4
	(0.001)	(0.011)	(**0.011**)	(0.013)	(0.013)
pendigi	0.474_5	0.014_2	$\mathbf{0.014_1}$	0.015_4	0.015_3
	(0.010)	(0.004)	(**0.004**)	(0.004)	(0.004)
scene-c	0.654_5	$\mathbf{0.386_1}$	0.389_2	0.394_4	0.394_3
	(0.006)	(**0.012**)	(0.013)	(0.012)	(0.011)
segment	0.241_5	0.009_4	$\mathbf{0.009_1}$	0.009_3	0.009_2
	(0.025)	(0.008)	(**0.007**)	(0.008)	(0.007)
shuttle	0.478_5	0.000_2	$\mathbf{0.000_1}$	0.000_3	0.000_4
	(0.020)	(0.001)	(**0.001**)	(0.001)	(0.001)
sonar	0.620_5	0.414_4	$\mathbf{0.412_1}$	0.413_3	0.413_2
	(0.009)	(0.073)	(**0.071**)	(0.071)	(0.071)
spambas	0.646_5	$\mathbf{0.178_1}$	0.178_2	0.182_4	0.182_3
	(0.004)	(**0.013**)	(0.013)	(0.014)	(0.014)
tic-tac	0.689_5	0.169_2	0.193_4	$\mathbf{0.169_1}$	0.170_3
	(0.000)	(0.031)	(0.034)	(**0.031**)	(0.030)
vehicle	0.575_5	0.070_2	$\mathbf{0.070_1}$	0.071_4	0.071_3
	(0.009)	(0.021)	(**0.020**)	(0.020)	(0.020)
vowel	0.422_5	$\mathbf{0.064_1}$	0.066_4	0.065_3	0.065_2
	(0.011)	(**0.019**)	(0.018)	(0.018)	(0.018)
wavefor	0.669_5	0.262_2	$\mathbf{0.261_1}$	0.266_3	0.266_4
	(0.004)	(0.011)	(**0.012**)	(0.012)	(0.012)
wdbc	0.422_5	0.090_2	$\mathbf{0.089_1}$	0.091_4	0.090_3
	(0.027)	(0.033)	(**0.032**)	(0.032)	(0.032)
wpbc	0.625_5	0.503_3	0.508_4	0.503_2	$\mathbf{0.502_1}$
	(0.012)	(0.036)	(0.044)	(0.033)	(**0.033**)
yeast	0.673_5	$\mathbf{0.540_1}$	0.548_4	0.541_2	0.541_3
	(0.002)	(**0.012**)	(0.015)	(0.011)	(0.011)
zoo	$\mathbf{0.006_1}$	0.013_3	0.013_4	0.013_5	0.013_2
	(**0.013**)	(0.020)	(0.020)	(0.020)	(0.020)
rank	4.76	1.96	2.06	3.24	2.98

Table 18: Brier score results for Ada-S (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results significant at $1.576e{-}13$, according to Friedman's test.

dataset	uncalibrated	beta	beta05	beta2	logistic
abalone	0.249_5	0.220_2	$\mathbf{0.220_1}$	0.222_3	0.222_4
	(0.000)	(0.003)	(**0.004**)	(0.003)	(0.003)
autos	0.165_5	0.074_2	$\mathbf{0.073_1}$	0.077_4	0.076_3
	(0.010)	(0.025)	(**0.024**)	(0.022)	(0.022)
balance	0.225_5	0.009_2	0.009_4	$\mathbf{0.009_1}$	0.009_3
	(0.001)	(0.006)	(0.006)	(**0.006**)	(0.006)
car	0.229_5	0.033_3	0.034_4	0.033_2	$\mathbf{0.033_1}$
	(0.001)	(0.006)	(0.006)	(0.006)	(**0.006**)
clevela	0.238_5	0.155_2	$\mathbf{0.155_1}$	0.156_4	0.156_3
	(0.002)	(0.018)	(**0.018**)	(0.017)	(0.017)
credit-	0.241_5	0.121_2	$\mathbf{0.121_1}$	0.124_4	0.124_3
	(0.001)	(0.013)	(**0.012**)	(0.013)	(0.013)
dermato	$\mathbf{0.006_1}$	0.006_4	0.006_2	0.007_5	0.006_3
	(**0.009**)	(0.009)	(0.009)	(0.009)	(0.009)
diabete	0.245_5	$\mathbf{0.178_1}$	0.179_2	0.184_4	0.184_3
	(0.001)	(**0.007**)	(0.009)	(0.007)	(0.007)
ecoli	0.125_5	0.037_4	0.037_2	0.037_3	$\mathbf{0.037_1}$
	(0.021)	(0.019)	(0.019)	(0.019)	(**0.019**)
flare	0.227_5	$\mathbf{0.133_1}$	0.139_4	0.133_2	0.133_3
	(0.006)	(**0.008**)	(0.010)	(0.008)	(0.008)
german	0.248_5	$\mathbf{0.184_1}$	0.186_2	0.190_3	0.190_4
	(0.000)	(**0.007**)	(0.009)	(0.006)	(0.006)
glass	0.233_5	$\mathbf{0.162_1}$	0.162_2	0.167_4	0.167_3
	(0.003)	(**0.019**)	(0.019)	(0.017)	(0.017)
heart-s	0.236_5	$\mathbf{0.154_1}$	0.155_2	0.155_4	0.155_3
	(0.002)	(**0.020**)	(0.020)	(0.019)	(0.019)
hepatit	0.186_5	0.124_3	0.125_4	0.123_2	$\mathbf{0.123_1}$
	(0.012)	(0.020)	(0.021)	(0.020)	(**0.020**)
horse	0.226_5	$\mathbf{0.133_1}$	0.137_4	0.136_3	0.136_2
	(0.003)	(**0.019**)	(0.018)	(0.018)	(0.018)
ionosph	0.171_5	0.054_2	$\mathbf{0.054_1}$	0.054_4	0.054_3
	(0.007)	(0.015)	(**0.016**)	(0.015)	(0.015)
iris	$\mathbf{0.000_1}$	0.000_3	0.000_2	0.000_5	0.000_4
	(**0.000**)	(0.000)	(0.000)	(0.000)	(0.000)
landsat	0.208_5	0.010_3	$\mathbf{0.010_1}$	0.010_2	0.010_4
	(0.003)	(0.002)	(**0.002**)	(0.002)	(0.002)
letter	0.193_5	0.006_2	$\mathbf{0.006_1}$	0.006_3	0.006_4
	(0.001)	(0.001)	(**0.001**)	(0.001)	(0.001)
libras-	0.052_5	$\mathbf{0.021_1}$	0.024_4	0.023_3	0.021_2
	(0.010)	(**0.013**)	(0.012)	(0.012)	(0.012)
lung-ca	$\mathbf{0.027_1}$	0.032_4	0.032_3	0.032_2	0.032_5
	(**0.024**)	(0.022)	(0.022)	(0.022)	(0.022)
mfeat-k	0.137_5	$\mathbf{0.006_1}$	0.006_4	0.006_3	0.006_2
	(0.009)	(**0.003**)	(0.003)	(0.003)	(0.003)
mfeat-m	0.009_5	0.002_3	$\mathbf{0.002_1}$	0.002_2	0.002_4
	(0.004)	(0.002)	(**0.002**)	(0.002)	(0.002)
mfeat-z	0.091_5	$\mathbf{0.007_1}$	0.007_2	0.007_4	0.007_3
	(0.011)	(**0.003**)	(0.003)	(0.003)	(0.003)
mushroo	0.175_5	0.000_3	$\mathbf{0.000_1}$	0.000_2	0.000_4
	(0.002)	(0.000)	(**0.000**)	(0.000)	(0.000)
optdigi	0.192_5	0.008_2	$\mathbf{0.008_1}$	0.008_4	0.008_3
	(0.005)	(0.002)	(**0.002**)	(0.002)	(0.002)
page-bl	0.235_5	0.024_2	$\mathbf{0.024_1}$	0.025_3	0.025_4
	(0.001)	(0.003)	(**0.003**)	(0.003)	(0.003)
pendigi	0.145_5	0.003_2	$\mathbf{0.003_1}$	0.003_3	0.003_4
	(0.005)	(0.001)	(**0.001**)	(0.001)	(0.001)
scene-c	0.231_5	$\mathbf{0.123_1}$	0.123_2	0.126_4	0.126_3
	(0.003)	(**0.005**)	(0.005)	(0.004)	(0.004)
segment	0.050_5	0.002_4	$\mathbf{0.002_1}$	0.002_3	0.002_2
	(0.008)	(0.001)	(**0.002**)	(0.001)	(0.001)
shuttle	0.146_5	$\mathbf{0.000_1}$	0.000_4	0.000_2	0.000_3
	(0.009)	(**0.000**)	(0.000)	(0.000)	(0.000)
sonar	0.214_5	0.131_2	$\mathbf{0.130_1}$	0.131_4	0.131_3
	(0.004)	(0.029)	(**0.029**)	(0.029)	(0.029)
spambas	0.227_5	$\mathbf{0.047_1}$	0.047_2	0.047_4	0.047_3
	(0.002)	(**0.003**)	(0.003)	(0.003)	(0.003)
tic-tac	0.248_5	0.045_2	0.056_4	$\mathbf{0.045_1}$	0.045_3
	(0.000)	(0.009)	(0.011)	(**0.009**)	(0.009)
vehicle	0.192_5	$\mathbf{0.018_1}$	0.018_2	0.018_3	0.018_4
	(0.004)	(**0.005**)	(0.005)	(0.005)	(0.005)
vowel	0.123_5	$\mathbf{0.017_1}$	0.018_4	0.017_3	0.017_2
	(0.004)	(**0.006**)	(0.005)	(0.005)	(0.005)
wavefor	0.238_5	0.080_2	$\mathbf{0.080_1}$	0.081_3	0.081_4
	(0.002)	(0.004)	(**0.004**)	(0.004)	(0.004)
wdbc	0.121_5	0.024_2	$\mathbf{0.024_1}$	0.024_4	0.024_3
	(0.011)	(0.010)	(**0.009**)	(0.009)	(0.009)
wpbc	0.216_5	0.164_3	0.166_4	0.164_2	$\mathbf{0.164_1}$
	(0.006)	(0.013)	(0.017)	(0.012)	(**0.012**)
yeast	0.240_5	$\mathbf{0.185_1}$	0.189_4	0.186_2	0.186_3
	(0.001)	(**0.004**)	(0.005)	(0.003)	(0.003)
zoo	$\mathbf{0.002_1}$	0.002_5	0.002_3	0.002_2	0.002_4
	(**0.003**)	(0.003)	(0.003)	(0.003)	(0.003)
rank	4.61	2.07	2.24	3.05	3.02

Table 19: Accuracy results for Ada-S in % (standard deviation in parentheses). Best results are marked in **bold** and subscript numbers indicate the ranks. Differences between results may occur after the third decimal digit. Results not significant, according to Friedman's test (p-value = 0.162).

dataset	uncalibrated	beta	beta[$m = 1/2$]	beta[a=b]	logistic
abalone	63.304_4	$\mathbf{63.778}_1$	63.091_5	63.696_2	63.696_3
	(1.430)	(**0.329**)	(1.404)	(0.313)	(0.308)
autos	91.495_5	91.609_2	91.550_4	91.555_3	$\mathbf{91.674}_1$
	(4.969)	(4.483)	(4.474)	(4.593)	(**4.618**)
balance	98.830_5	98.975_2	98.895_4	98.959_3	$\mathbf{99.023}_1$
	(1.015)	(1.097)	(0.927)	(1.135)	(**1.064**)
car	95.034_5	$95.156_{2.5}$	95.052_4	$95.156_{2.5}$	$\mathbf{95.156}_1$
	(1.059)	(1.101)	(1.081)	(1.101)	(**1.111**)
clevela	$\mathbf{79.483}_1$	78.227_5	78.772_2	$78.363_{3.5}$	$78.363_{3.5}$
	(**4.701**)	(5.020)	(4.725)	(5.046)	(5.046)
credit-	84.258_2	83.968_3	$\mathbf{84.396}_1$	$83.877_{4.5}$	$83.877_{4.5}$
	(2.464)	(2.696)	(**2.350**)	(2.735)	(2.735)
dermato	$99.408_{2.5}$	$99.408_{2.5}$	$99.408_{2.5}$	99.351_5	$99.408_{2.5}$
	(1.111)	(1.111)	(1.111)	(1.153)	(1.111)
diabete	74.115_2	73.591_3	$\mathbf{74.245}_1$	72.096_4	72.070_5
	(2.873)	(2.166)	(**3.130**)	(2.632)	(2.658)
ecoli	$\mathbf{96.083}_1$	95.725_5	95.729_4	95.754_3	95.783_2
	(**2.450**)	(2.574)	(2.572)	(2.566)	(2.523)
flare	82.145_4	82.260_2	82.145_5	82.260_2	82.260_2
	(1.267)	(1.135)	(1.251)	(1.135)	(1.135)
german	73.810_2	71.530_3	$\mathbf{73.860}_1$	70.600_4	70.560_5
	(2.470)	(2.088)	(**2.458**)	(1.161)	(1.155)
glass	$\mathbf{80.460}_1$	79.168_3	79.978_2	78.421_5	78.516_4
	(**4.919**)	(5.608)	(4.808)	(4.343)	(4.386)
heart-s	$\mathbf{79.074}_1$	78.222_5	78.593_2	$78.333_{3.5}$	$78.333_{3.5}$
	(**5.347**)	(5.083)	(5.526)	(5.360)	(5.360)
hepatit	$\mathbf{82.916}_1$	80.733_5	81.698_3	81.826_2	81.697_4
	(**6.088**)	(4.211)	(5.907)	(4.429)	(4.561)
horse	$\mathbf{83.870}_1$	83.038_5	83.672_2	83.339_4	83.373_3
	(**4.325**)	(3.929)	(4.381)	(3.827)	(3.849)
ionosph	$\mathbf{93.336}_1$	93.223_5	93.252_4	$93.336_{2.5}$	$93.336_{2.5}$
	(**2.652**)	(2.511)	(2.543)	(2.392)	(2.392)
iris	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
landsat	98.636_5	98.645_4	$\mathbf{98.656}_1$	98.650_3	98.651_2
	(0.343)	(0.336)	(**0.343**)	(0.348)	(0.347)
letter	99.202_2	99.200_3	$\mathbf{99.207}_1$	99.200_4	99.198_5
	(0.099)	(0.102)	(**0.097**)	(0.101)	(0.102)
libras-	97.080_5	$\mathbf{97.693}_1$	97.307_4	97.553_3	97.640_2
	(1.580)	(**1.809**)	(1.620)	(1.759)	(1.811)
lung-ca	$\mathbf{97.295}_1$	95.621_4	$95.726_{2.5}$	$95.726_{2.5}$	95.521_5
	(**3.820**)	(4.386)	(4.297)	(4.297)	(4.457)
mfeat-k	$\mathbf{99.330}_1$	99.325_2	99.315_3	$99.315_{4.5}$	$99.315_{4.5}$
	(**0.393**)	(0.347)	(0.353)	(0.349)	(0.349)
mfeat-m	99.765_5	99.775_2	$\mathbf{99.775}_1$	$99.770_{3.5}$	$99.770_{3.5}$
	(0.211)	(0.216)	(**0.210**)	(0.213)	(0.213)
mfeat-z	99.155_5	99.185_3	99.170_4	99.190_2	$\mathbf{99.195}_1$
	(0.413)	(0.413)	(0.409)	(0.403)	(**0.402**)
mushroo	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
optdigi	98.899_5	$\mathbf{98.925}_1$	98.925_2	$98.918_{3.5}$	$98.918_{3.5}$
	(0.289)	(**0.281**)	(0.280)	(0.289)	(0.289)
page-bl	$\mathbf{96.996}_1$	96.938_3	96.962_2	96.865_5	96.867_4
	(**0.461**)	(0.426)	(0.481)	(0.468)	(0.470)
pendigi	99.674_2	99.671_3	$\mathbf{99.685}_1$	99.662_4	99.660_5
	(0.115)	(0.118)	(**0.108**)	(0.121)	(0.122)
scene-c	$\mathbf{83.224}_1$	82.526_3	83.120_2	81.508_4	81.495_5
	(**1.450**)	(1.168)	(1.592)	(1.108)	(1.105)
segment	99.732_5	99.740_4	$\mathbf{99.779}_1$	99.758_2	99.753_3
	(0.250)	(0.219)	(**0.207**)	(0.199)	(0.205)
shuttle	99.995_5	99.998_3	$\mathbf{99.998}_1$	99.998_4	99.998_2
	(0.006)	(0.005)	(**0.005**)	(0.005)	(0.005)
sonar	81.865_3	81.893_2	$\mathbf{82.096}_1$	81.802_5	81.851_4
	(6.839)	(6.172)	(**6.162**)	(6.497)	(6.459)
spambas	94.256_5	$\mathbf{94.523}_1$	94.508_4	$94.508_{2.5}$	$94.508_{2.5}$
	(0.637)	(**0.618**)	(0.560)	(0.573)	(0.573)
tic-tac	91.618_5	95.189_2	91.702_4	$\mathbf{95.199}_1$	95.178_3
	(2.118)	(1.469)	(2.042)	(**1.479**)	(1.458)
vehicle	97.766_5	97.849_4	$\mathbf{97.883}_1$	$97.861_{2.5}$	$97.861_{2.5}$
	(0.941)	(1.009)	(**0.979**)	(0.967)	(0.967)
vowel	97.455_5	$\mathbf{97.960}_1$	97.687_4	97.889_3	97.909_2
	(0.866)	(**1.030**)	(0.896)	(0.972)	(0.973)
wavefor	88.830_5	88.986_3	88.962_4	$\mathbf{88.996}_1$	88.996_2
	(0.788)	(0.773)	(0.774)	(**0.820**)	(0.827)
wdbc	$\mathbf{97.082}_1$	97.011_3	97.012_2	96.940_5	96.958_4
	(**1.437**)	(1.668)	(1.712)	(1.643)	(1.643)
wpbc	77.806_2	76.605_3	$\mathbf{78.416}_1$	$76.457_{4.5}$	$76.457_{4.5}$
	(4.605)	(2.390)	(**4.369**)	(2.117)	(2.117)
yeast	72.639_2	69.177_5	$\mathbf{72.896}_1$	69.264_3	69.251_4
	(2.076)	(0.654)	(**2.101**)	(0.810)	(0.814)
zoo	100.000_3	100.000_3	100.000_3	100.000_3	100.000_3
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
rank	3.01	3.00	2.51	3.29	3.18

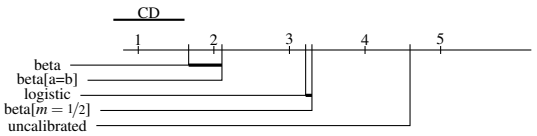

Figure 5: Critical difference diagram for Brier scores results of parametric methods with Naive Bayes as base classifier. Friedman test shows significance at p-value=$2.368e{-}18$.

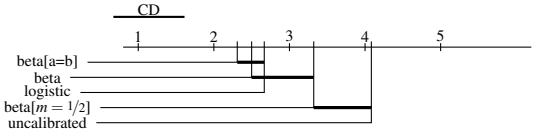

Figure 6: Critical difference diagram for accuracy results of parametric methods with Naive Bayes as base classifier. Friedman test shows significance at p-value=$3.844e{-}08$.

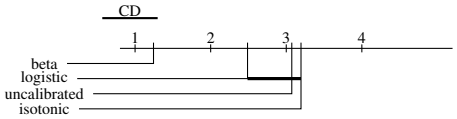

Figure 7: Critical difference diagram for log-loss results with Ada-O as base classifier. Friedman test shows significance at p-value=$1.008e{-}12$.

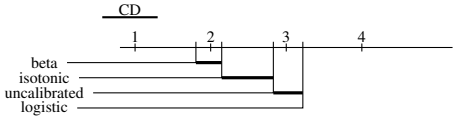

Figure 8: Critical difference diagram for Brier score results with Ada-O as base classifier. Friedman test shows significance at p-value=$3.745e{-}06$.

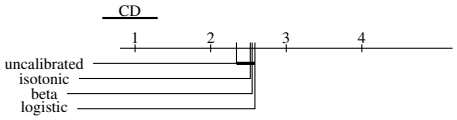

Figure 9: Critical difference diagram for accuracy results with Ada-O as base classifier. Results not significant, according to Friedman's test (p-value = 0.813).

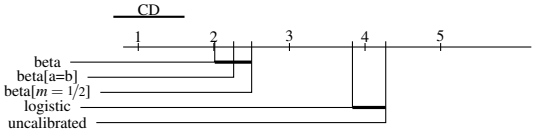

Figure 10: Critical difference diagram for log-loss results of parametric methods with Ada-O as base classifier. Friedman test shows significance at p-value=$1.598e{-}15$.

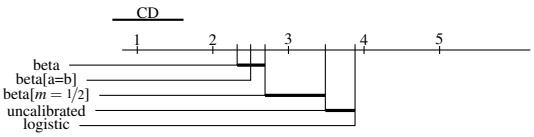

Figure 11: Critical difference diagram for Brier score results of parametric methods with Ada-O as base classifier. Friedman test shows significance at p-value=$1.400e-06$.

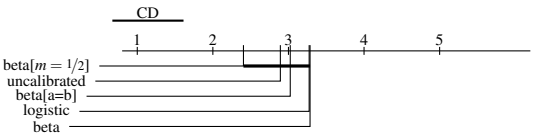

Figure 12: Critical difference diagram for Brier score results of parametric methods with Ada-O as base classifier. Results not significant, according to Friedman's test (p-value = 0.048).

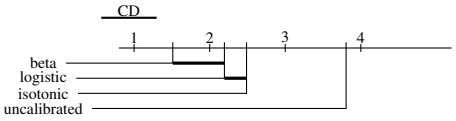

Figure 13: Critical difference diagram for log-loss results with Ada-S as base classifier. Friedman test shows significance at p-value=$4.733e-15$.

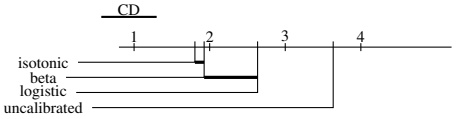

Figure 14: Critical difference diagram for Brier score results with Ada-S as base classifier. Friedman test shows significance at p-value=$5.836e-13$.

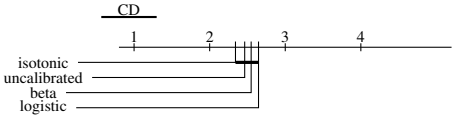

Figure 15: Critical difference diagram for accuracy results with Ada-S as base classifier. Results not significant, according to Friedman's test (p-value = 0.660).

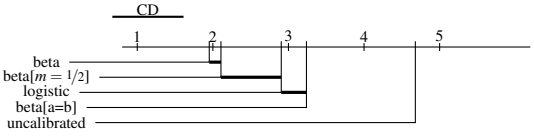

Figure 16: Critical difference diagram for log-loss results of parametric methods with Ada-S as base classifier. Friedman test shows significance at p-value=$2.799e-17$.

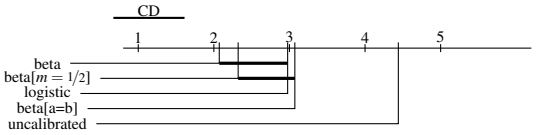

Figure 17: Critical difference diagram for Brier score results of parametric methods with Ada-S as base classifier. Friedman test shows significance at p-value=$1.576e-13$.

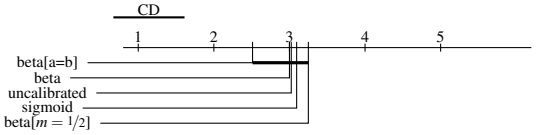

Figure 18: Critical difference diagram for accuracy results of parametric methods with Ada-S as base classifier. Results not significant, according to Friedman's test (p-value = 0.162).

References

J. Friedman, T. Hastie, R. Tibshirani, et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2):337–407, 2000.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *J. Machine Learning Research*, 12:2825–2830, 2011.

J. Platt. Probabilities for SV machines. In A. Smola, P. Bartlett, B. Schölkopf, and D. Schuurmans, editors, *Adv. Large Margin Classifiers*, pages 61–74. MIT Press, 2000.

J. Zhu, H. Zou, S. Rosset, and T. Hastie. Multi-class AdaBoost. *Statistics and its Interface*, 2(3):349–360, 2009.