Output tables for the test of Multiple comparisons.

January 10, 2022

1 Average rankings of Friedman test

Average ranks obtained by applying the Friedman procedure

Friedman statistic considering reduction performance (distributed according to chi-square with 12 degrees of freedom: 73.811919.

P-value computed by Friedman Test: 1.1092249341260185E-10.

Algorithm	Ranking
best-precision	8.5192
best-recall	5.5192
balanced	3.8846
promethee-precision	8.5192
promethee-recall	5.5192
bac	6.8269
precision	9.6346
recall	10.3269
f1	5.25
auc	6.3654
gmean	7.0577
AdaBoost	7.9615
Bagging	5.6154

Table 1: Average Rankings of the algorithms

2 Post hoc comparisons

Results achieved on post hoc comparisons for $\alpha = 0.05$, $\alpha = 0.10$ and adjusted p-values.

2.1 P-values for $\alpha = 0.05$

Nemenyi's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.000641.

i	algorithms	$z = (R_0 - R_i)/SE$	p
78	balanced vs. recall	5.964418	0
77	balanced vs. precision	5.323466	0
76	recall vs. f1	4.700317	0.000003
75	best-recall vs. recall	4.451058	0.000009
74	promethee-recall vs. recall	4.451058	0.000009
73	recall vs. Bagging	4.362037	0.000013
72	best-precision vs. balanced	4.29082	0.000018
71	balanced vs. promethee-precision	4.29082	0.000018
70	precision vs. f1	4.059365	0.000049
69	best-recall vs. precision	3.810106	0.000139
68	promethee-recall vs. precision	3.810106	0.000139
67	balanced vs. AdaBoost	3.774497	0.00016
66	precision vs. Bagging	3.721085	0.000198
65	recall vs. auc	3.667672	0.000245
64	bac vs. recall	3.24037	0.001194
63	precision vs. auc	3.02672	0.002472
62	best-precision vs. f1	3.02672	0.002472
61	promethee-precision vs. f1	3.02672	0.002472
60	recall vs. gmean	3.02672	0.002472
59	balanced vs. gmean	2.937698	0.003307
58	best-precision vs. best-recall	2.77746	0.005479
57	best-precision vs. promethee-recall	2.77746	0.005479
56	best-recall vs. promethee-precision	2.77746	0.005479
55	promethee-precision vs. promethee-recall	2.77746	0.005479
54	balanced vs. bac	2.724048	0.006449
53	best-precision vs. Bagging	2.688439	0.007179
52	promethee-precision vs. Bagging	2.688439	0.007179
51	bac vs. precision	2.599418	0.009338
50	f1 vs. AdaBoost	2.510397	0.01206
49	precision vs. gmean	2.385767	0.017044
48	balanced vs. auc	2.296746	0.021633
47	best-recall vs. AdaBoost	2.261138	0.023751
46	promethee-recall vs. AdaBoost	2.261138	0.023751
45	recall vs. AdaBoost	2.189921	0.02853
44	AdaBoost vs. Bagging	2.172116	0.029847
43	best-precision vs. auc	1.994074	0.046144
42	promethee-precision vs. auc	1.994074	0.046144
41	best-precision vs. recall	1.673598	0.09421
40	promethee-precision vs. recall	1.673598	0.09421
39	f1 vs. gmean	1.673598	0.09421
38	balanced vs. Bagging	1.602381	0.109071
37	best-precision vs. bac	1.566772	0.117168
36	promethee-precision vs. bac	1.566772	0.117168
35	precision vs. AdaBoost	1.548968	0.121389
34	best-recall vs. balanced	1.51336	0.130188
33	balanced vs. promethee-recall	1.51336	0.130188
32	auc vs. AdaBoost	1.477751	0.139474
31	bac vs. f1	1.459947	0.144305
30	best-recall vs. gmean	1.424339	0.154348
29	promethee-recall vs. gmean	1.424339	0.154348
28	best-precision vs. gmean	1.353122	0.176017
27	promethee-precision vs. gmean	1.353122	0.176017
26	gmean vs. Bagging	1.335317	0.181772
25	balanced vs. f1	1.264101	0.206194
24	best-recall vs. bac	1.210688	0.226015
23	promethee-recall vs. bac	1.210688	0.226015
22	bac vs. Bagging	1.121667	0.262004
21	bac vs. AdaBoost	1.05045	0.293511
20	best-precision vs. precision	1.032645	0.30177
19	promethee-precision vs. precision	1.032645	0.30177
18	f1 vs. auc	1.032645	0.30177
17	gmean vs. AdaBoost	0.836799	0.402706
16	best-recall vs. auc	0.783386	0.4334
15	promethee-recall vs. auc	0.783386	0.4334
14	auc vs. Bagging	0.694365	0.487453
13	auc vs. gmean	0.640952	0.521554
12	precision vs. recall	0.640952	0.521554
11	best-precision vs. AdaBoost	0.516323	0.605629
10	promethee-precision vs. AdaBoost	0.516323	0.605629
9	bac vs. auc	0.427302	0.66916
8	f1 vs. Bagging	0.33828	0.735152
7	best-recall vs. f1	0.249259	0.80316
6	promethee-recall vs. f1	0.249259	0.80316
5	bac vs. gmean	0.213651	0.830819
4	best-recall vs. Bagging	0.089021	0.929065
3	promethee-recall vs. Bagging	0.089021	0.929065
2	best-precision vs. promethee-precision	0	1
1	best-recall vs. promethee-recall	0	1

Table 2: P-values Table for $\alpha = 0.05$

2.2 P-values for $\alpha = 0.10$

Nemenyi's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.001282.

i	algorithms	$z = (R_0 - R_i)/SE$	p
78	balanced vs. recall	5.964418	0
77	balanced vs. precision	5.323466	0
76	recall vs. f1	4.700317	0.000003
75	best-recall vs. recall	4.451058	0.000009
74	promethee-recall vs. recall	4.451058	0.000009
73	recall vs. Bagging	4.362037	0.000013
72	best-precision vs. balanced	4.29082	0.000018
71	balanced vs. promethee-precision	4.29082	0.000018
70	precision vs. f1	4.059365	0.000049
69	best-recall vs. precision	3.810106	0.000139
68	promethee-recall vs. precision	3.810106	0.000139
67	balanced vs. AdaBoost	3.774497	0.00016
66	precision vs. Bagging	3.721085	0.000198
65	recall vs. auc	3.667672	0.000245
64	bac vs. recall	3.24037	0.001194
63	precision vs. auc	3.02672	0.002472
62	best-precision vs. f1	3.02672	0.002472
61	promethee-precision vs. f1	3.02672	0.002472
60	recall vs. gmean	3.02672	0.002472
59	balanced vs. gmean	2.937698	0.003307
58	best-precision vs. best-recall	2.77746	0.005479
57	best-precision vs. promethee-recall	2.77746	0.005479
56	best-recall vs. promethee-precision	2.77746	0.005479
55	promethee-precision vs. promethee-recall	2.77746	0.005479
54	balanced vs. bac	2.724048	0.006449
53	best-precision vs. Bagging	2.688439	0.007179
52	promethee-precision vs. Bagging	2.688439	0.007179
51	bac vs. precision	2.599418	0.009338
50	f1 vs. AdaBoost	2.510397	0.01206
49	precision vs. gmean	2.385767	0.017044
48	balanced vs. auc	2.296746	0.021633
47	best-recall vs. AdaBoost	2.261138	0.023751
46	promethee-recall vs. AdaBoost	2.261138	0.023751
45	recall vs. AdaBoost	2.189921	0.02853
44	AdaBoost vs. Bagging	2.172116	0.029847
43	best-precision vs. auc	1.994074	0.046144
42	promethee-precision vs. auc	1.994074	0.046144
41	best-precision vs. recall	1.673598	0.09421
40	promethee-precision vs. recall	1.673598	0.09421
39	f1 vs. gmean	1.673598	0.09421
38	balanced vs. Bagging	1.602381	0.109071
37	best-precision vs. bac	1.566772	0.117168
36	promethee-precision vs. bac	1.566772	0.117168
35	precision vs. AdaBoost	1.548968	0.121389
34	best-recall vs. balanced	1.51336	0.130188
33	balanced vs. promethee-recall	1.51336	0.130188
32	auc vs. AdaBoost	1.477751	0.139474
31	bac vs. f1	1.459947	0.144305
30	best-recall vs. gmean	1.424339	0.154348
29	promethee-recall vs. gmean	1.424339	0.154348
28	best-precision vs. gmean	1.353122	0.176017
27	promethee-precision vs. gmean	1.353122	0.176017
26	gmean vs. Bagging	1.335317	0.181772
25	balanced vs. f1	1.264101	0.206194
24	best-recall vs. bac	1.210688	0.226015
23	promethee-recall vs. bac	1.210688	0.226015
22	bac vs. Bagging	1.121667	0.262004
21	bac vs. AdaBoost	1.05045	0.293511
20	best-precision vs. precision	1.032645	0.30177
19	promethee-precision vs. precision	1.032645	0.30177
18	f1 vs. auc	1.032645	0.30177
17	gmean vs. AdaBoost	0.836799	0.402706
16	best-recall vs. auc	0.783386	0.4334
15	promethee-recall vs. auc	0.783386	0.4334
14	auc vs. Bagging	0.694365	0.487453
13	auc vs. gmean	0.640952	0.521554
12	precision vs. recall	0.640952	0.521554
11	best-precision vs. AdaBoost	0.516323	0.605629
10	promethee-precision vs. AdaBoost	0.516323	0.605629
9	bac vs. auc	0.427302	0.66916
8	f1 vs. Bagging	0.33828	0.735152
7	best-recall vs. f1	0.249259	0.80316
6	promethee-recall vs. f1	0.249259	0.80316
5	bac vs. gmean	0.213651	0.830819
4	best-recall vs. Bagging	0.089021	0.929065
3	promethee-recall vs. Bagging	0.089021	0.929065
2	best-precision vs. promethee-precision	0	1
1	best-recall vs. promethee-recall	0	1

Table 3: P-values Table for $\alpha = 0.10$

2.3 Adjusted p-values

i	hypothesis	unadjusted p	p_{Neme}
1	balanced vs . recall	0	0
2	balanced vs . precision	0	0.000008
3	recall vs . f1	0.000003	0.000203
4	best-recall vs . recall	0.000009	0.000666
5	promethee-recall vs . recall	0.000009	0.000666
6	recall vs .Bagging	0.000013	0.001005
7	best-precision vs . balanced	0.000018	0.001389
8	balanced vs . promethee-precision	0.000018	0.001389
9	precision vs . f1	0.000049	0.003838
10	best-recall vs . precision	0.000139	0.010835
11	promethee-recall vs . precision	0.000139	0.010835
12	balanced vs .AdaBoost	0.00016	0.012506
13	precision vs .Bagging	0.000198	0.015473
14	recall vs . auc	0.000245	0.019092
15	bac vs . recall	0.001194	0.093112
16	precision vs . auc	0.002472	0.192834
17	best-precision vs . f1	0.002472	0.192834
18	promethee-precision vs . f1	0.002472	0.192834
19	recall vs . gmean	0.002472	0.192834
20	balanced vs . gmean	0.003307	0.257914
21	best-precision vs . best-recall	0.005479	0.427327
22	best-precision vs . promethee-recall	0.005479	0.427327
23	best-recall vs . promethee-precision	0.005479	0.427327
24	promethee-precision vs . promethee-recall	0.005479	0.427327
25	balanced vs . bac	0.006449	0.503
26	best-precision vs .Bagging	0.007179	0.559938
27	promethee-precision vs .Bagging	0.007179	0.559938
28	bac vs . precision	0.009338	0.72838
29	f1 vs .AdaBoost	0.01206	0.940645
30	precision vs . gmean	0.017044	1.329396
31	balanced vs . auc	0.021633	1.687394
32	best-recall vs .AdaBoost	0.023751	1.852558
33	promethee-recall vs .AdaBoost	0.023751	1.852558
34	recall vs .AdaBoost	0.02853	2.22534
35	AdaBoost vs .Bagging	0.029847	2.328057
36	best-precision vs . auc	0.046144	3.599229
37	promethee-precision vs . auc	0.046144	3.599229
38	best-precision vs . recall	0.09421	7.348353
39	promethee-precision vs . recall	0.09421	7.348353
40	f1 vs . gmean	0.09421	7.348353
41	balanced vs .Bagging	0.109071	8.507569
42	best-precision vs . bac	0.117168	9.139095
43	promethee-precision vs . bac	0.117168	9.139095
44	precision vs .AdaBoost	0.121389	9.46837
45	best-recall vs . balanced	0.130188	10.154687
46	balanced vs . promethee-recall	0.130188	10.154687
47	auc vs .AdaBoost	0.139474	10.879
48	bac vs . f1	0.144305	11.25576
49	best-recall vs . gmean	0.154348	12.039181
50	promethee-recall vs . gmean	0.154348	12.039181
51	best-precision vs . gmean	0.176017	13.729307
52	promethee-precision vs . gmean	0.176017	13.729307
53	gmean vs .Bagging	0.181772	14.178253
54	balanced vs . f1	0.206194	16.083128
55	best-recall vs . bac	0.226015	17.629176
56	promethee-recall vs . bac	0.226015	17.629176
57	bac vs .Bagging	0.262004	20.436327
58	bac vs .AdaBoost	0.293511	22.893889
59	best-precision vs . precision	0.30177	23.538047
60	promethee-precision vs . precision	0.30177	23.538047
61	f1 vs . auc	0.30177	23.538047
62	gmean vs .AdaBoost	0.402706	31.411036
63	best-recall vs . auc	0.4334	33.805226
64	promethee-recall vs . auc	0.4334	33.805226
65	auc vs .Bagging	0.487453	38.021357
66	auc vs . gmean	0.521554	40.681183
67	precision vs . recall	0.521554	40.681183
68	best-precision vs .AdaBoost	0.605629	47.239063
69	promethee-precision vs .AdaBoost	0.605629	47.239063
70	bac vs . auc	0.66916	52.194455
71	f1 vs .Bagging	0.735152	57.341846
72	best-recall vs . f1	0.80316	62.646499
73	promethee-recall vs . f1	0.80316	62.646499
74	bac vs . gmean	0.830819	64.803912
75	best-recall vs .Bagging	0.929065	72.467077
76	promethee-recall vs .Bagging	0.929065	72.467077
77	best-precision vs . promethee-precision	1	78
78	best-recall vs . promethee-recall	1	78

Table 4: Adjusted p-values