Results

May 7, 2024

1 Tables of Friedman, Bonferroni-Dunn, Holm, Hochberg and Hommel Tests

Table 1: Average Rankings of the algorithms

Algorithm	Ranking
LWRK	3.42
MS	5.10
SPT	2.17
WINQ	5.45
—(3) AS—model—AS-model-0	2.71
—(0) NN—Baseline-model-0	6.50
—(3) AS—model—AS more rules-model-1	2.64

Friedman statistic considering reduction performance (distributed according to chi-square with 6 degrees of freedom: 1077.8853571428447.
P-value computed by Friedman Test: 0.0.

Iman and Davenport statistic considering reduction performance (distributed according to F-distribution with 6 and 1794 degrees of freedom: 446.3110185808318.
P-value computed by Iman and Daveport Test: 0.0.

Table 2: Holm / Hochberg Table for $\alpha = 0.05$

i	algorithm	$z = (R_0 - R_i)/SE$	p	Holm/Hochberg/Hommel
6	—(0) NN—Baseline-model-0	24.54879252194937	4.4553413308992614E-133	0.008333333333333333
5	WINQ	18.59585207205392	3.471524840767506E-77	0.01
4	MS	16.573742141454566	1.0789706243090323E-61	0.0125
3	LWRK	7.09628298074823	1.281569344021574E-12	0.016666666666666666
2	—(3) AS—model—AS-model-0	3.0426140077242385	0.0023453295675642475	0.025
1	—(3) AS—model—AS more rules-model-1	2.636302199239389	0.008381504991182207	0.05

Bonferroni-Dunn's procedure rejects those hypotheses that have a p-value $\leq 0.008333333333333333$.
Hochberg's procedure rejects those hypotheses that have a p-value ≤ 0.05.
Hommel's procedure rejects all hypotheses.

Table 3: Holm / Hochberg Table for $\alpha = 0.10$

i	algorithm	$z = (R_0 - R_i)/SE$	p	Holm/Hochberg/Hommel
6	—(0) NN—Baseline-model-0	24.54879252194937	4.4553413308992614E-133	0.016666666666666666
5	WINQ	18.59585207205392	3.471524840767506E-77	0.02
4	MS	16.573742141454566	1.0789706243090323E-61	0.025
3	LWRK	7.09628298074823	1.281569344021574E-12	0.03333333333333333
2	—(3) AS—model—AS-model-0	3.0426140077242385	0.0023453295675642475	0.05
1	—(3) AS—model—AS more rules-model-1	2.636302199239389	0.008381504991182207	0.1

Bonferroni-Dunn's procedure rejects those hypotheses that have a p-value $\leq 0.016666666666666666$.
Hochberg's procedure rejects those hypotheses that have a p-value ≤ 0.1.
Hommel's procedure rejects all hypotheses.

Table 4: Adjusted p-values

i	algorithm	unadjusted p	P_{Bonf}	P_{Holm}	P_{Hoch}	P_{Homm}
1	—(0) NN—Baseline-model-0	4.4553413308992614E-133	2.6732047985395567E-132	2.6732047985395567E-132	2.6732047985395567E-132	2.6732047985395567E-132
2	WINQ	3.471524840767506E-77	2.0829149044605037E-76	1.735762420383753E-76	1.735762420383753E-76	1.735762420383753E-76
3	MS	1.0789706243090323E-61	6.473823745854194E-61	4.3158824972361294E-61	4.3158824972361294E-61	4.3158824972361294E-61
4	LWRK	1.281569344021574E-12	7.689416064129444E-12	3.844708032064722E-12	3.844708032064722E-12	3.844708032064722E-12
5	—(3) AS—model—AS-model-0	0.0023453295675642475	0.014071977405385485	0.004690659135128495	0.004690659135128495	0.004690659135128495
6	—(3) AS—model—AS more rules-model-1	0.008381504991182207	0.05028902994709324	0.008381504991182207	0.008381504991182207	0.008381504991182207

Table 5: Holm / Shaffer Table for $\alpha = 0.05$

i	algorithms	$z = (R_0 - R_i)/SE$	p	Holm	Shaffer
21	SPT vs. —(0) NN—Baseline-model-0	24.54879252194937	4.4553413308992614E-133	0.002380952380952381	0.002380952380952381
20	—(0) NN—Baseline-model-0 vs. —(3) AS—model—AS more rules-model-1	21.912490322709978	1.9748327423595857E-106	0.0025	0.003333333333333333
19	—(3) AS—model—AS-model-0 vs. —(0) NN—Baseline-model-0	21.506178514225127	1.3627983531605862E-102	0.002631578947368421	0.003333333333333333
18	SPT vs. WINQ	18.59585207205392	3.471524840767506E-77	0.002777777777777778	0.003333333333333333
17	LWRK vs. —(0) NN—Baseline-model-0	17.452509541201135	3.294354664957412E-68	0.002941176470588253	0.003333333333333333
16	MS vs. SPT	16.573742141454566	1.0789706243090323E-61	0.003125	0.003333333333333333
15	WINQ vs. —(3) AS—model—AS more rules-model-1	15.959549872814527	2.4448828830165487E-57	0.003333333333333335	0.003333333333333335
14	WINQ vs. —(3) AS—model—AS-model-0	15.553238064329678	1.5126085316274575E-54	0.0035714285714285718	0.004545454545454546
13	MS vs. —(3) AS—model—AS more rules-model-1	13.937439942215178	3.751570720872638E-44	0.0038461538461538464	0.004545454545454546
12	MS vs. —(3) AS—model—AS-model-0	13.53112813373033	1.0243940276055593E-41	0.004166666666666667	0.004545454545454546
11	LWRK vs. WINQ	11.499569091305688	1.325756343419097E-30	0.004545454545454546	0.004545454545454546
10	LWRK vs. MS	9.477459160706339	2.605499077861517E-21	0.005	0.005
9	MS vs. —(0) NN—Baseline-model-0	7.975050380494798	1.523186734970824E-15	0.005555555555555556	0.005555555555555556
8	LWRK vs. SPT	7.09628298074823	1.281569344021574E-12	0.00625	0.0071428571428571435
7	WINQ vs. —(0) NN—Baseline-model-0	5.952940449895449	2.6336714755298126E-9	0.0071428571428571435	0.0071428571428571435
6	LWRK vs. —(3) AS—model—AS more rules-model-1	4.459980781508841	8.196700267193257E-6	0.008333333333333333	0.008333333333333333
5	LWRK vs. —(3) AS—model—AS-model-0	4.053668973023991	5.042055036357275E-5	0.01	0.01
4	SPT vs. —(3) AS—model—AS-model-0	3.0426140077242385	0.0023453295675642475	0.0125	0.0125
3	SPT vs. —(3) AS—model—AS more rules-model-1	2.636302199239389	0.008381504991182207	0.016666666666666666	0.016666666666666666
2	MS vs. WINQ	2.0221099305993495	0.04316499633141476	0.025	0.025
1	—(3) AS—model—AS-model-0 vs. —(3) AS—model—AS more rules-model-1	0.4063118084848493	0.6845135100551373	0.05	0.05

Nemenyi's procedure rejects those hypotheses that have a p-value $\leq 0.002380952380952381$.
Holm's procedure rejects those hypotheses that have a p-value ≤ 0.025.
Shaffer's procedure rejects those hypotheses that have a p-value $\leq 0.002380952380952381$.
Bergmann's procedure rejects these hypotheses:

- LWRK vs. MS

- LWRK vs. SPT

- LWRK vs. WINQ

- LWRK vs. —(3) AS—model—AS-model-0

- LWRK vs. —(0) NN—Baseline-model-0

- LWRK vs. —(3) AS—model—AS more rules-model-1

- MS vs. SPT

- MS vs. —(3) AS—model—AS-model-0

- MS vs. —(0) NN—Baseline-model-0

- MS vs. —(3) AS—model—AS more rules-model-1

- SPT vs. WINQ

- SPT vs. —(3) AS—model—AS-model-0

- SPT vs. —(0) NN—Baseline-model-0

- SPT vs. —(3) AS—model—AS more rules-model-1

- WINQ vs. —(3) AS—model—AS-model-0

- WINQ vs. —(0) NN—Baseline-model-0

- WINQ vs. —(3) AS—model—AS more rules-model-1

- —(3) AS—model—AS-model-0 vs. —(0) NN—Baseline-model-0

- —(0) NN—Baseline-model-0 vs. —(3) AS—model—AS more rules-model-1

Table 6: Holm / Shaffer Table for $\alpha = 0.10$

i	algorithms	$z = (R_0 - R_i)/SE$	p	Holm	Shaffer
21	SPT vs. —(0) NN—Baseline-model-0	24.54879252194937	4.4553413308992614E-133	0.004761904761904762	0.004761904761904762
20	—(0) NN—Baseline-model-0 vs. —(3) AS—model—AS more rules-model-1	21.912490322709978	1.9748327423595857E-106	0.005	0.006666666666666667
19	—(3) AS—model—AS-model-0 vs. —(0) NN—Baseline-model-0	21.506178514225127	1.3627983531605862E-102	0.00526315789473684	0.006666666666666667
18	SPT vs. WINQ	18.59585207205392	3.471524840767506E-77	0.005555555555555556	0.006666666666666667
17	LWRK vs. —(0) NN—Baseline-model-0	17.452509541201135	3.294354664957412E-68	0.0058823529411764705	0.006666666666666667
16	MS vs. SPT	16.573742141454566	1.0789706243090323E-61	0.00625	0.006666666666666667
15	WINQ vs. —(3) AS—model—AS more rules-model-1	15.959549872814527	2.4448828830165487E-57	0.006666666666666667	0.006666666666666667
14	WINQ vs. —(3) AS—model—AS-model-0	15.553238064329678	1.5126085316274575E-54	0.0071428571428571435	0.00909090909090909092
13	MS vs. —(3) AS—model—AS more rules-model-1	13.937439942215178	3.751570720872638E-44	0.007692307692307693	0.00909090909090909092
12	MS vs. —(3) AS—model—AS-model-0	13.531128133730329	1.024394027605593E-41	0.008333333333333333	0.00909090909090909092
11	LWRK vs. WINQ	11.499569091305688	1.3257563434190097E-30	0.00909090909090909092	0.00909090909090909092
10	LWRK vs. MS	9.477459160706339	2.6054990778615172E-21	0.01	0.01
9	MS vs. —(0) NN—Baseline-model-0	7.975050380494798	1.523186734970874E08248E-15	0.011111111111111112	0.011111111111111112
8	LWRK vs. SPT	7.09628298074823	1.281569344021574E-12	0.0125	0.014285714285714287
7	WINQ vs. —(0) NN—Baseline-model-0	5.952940449895449	2.6336714755298126E-9	0.014285714285714287	0.014285714285714287
6	LWRK vs. —(3) AS—model—AS more rules-model-1	4.459980781508841	8.196700267193257E-6	0.016666666666666666	0.016666666666666666
5	LWRK vs. —(3) AS—model—AS-model-0	4.053668973023991	5.042055036357275E-5	0.02	0.02
4	SPT vs. —(3) AS—model—AS-model-0	3.042614077242385	0.0023453295675642475	0.025	0.025
3	SPT vs. —(3) AS—model—AS more rules-model-1	2.6363021992393389	0.008381504991182207	0.03333333333333333	0.03333333333333333
2	MS vs. WINQ	2.0221099305993495	0.04316499633141476	0.05	0.05
1	—(3) AS—model—AS-model-0 vs. —(3) AS—model—AS more rules-model-1	0.4063118084848493	0.6845135100551373	0.1	0.1

Nemenyi's procedure rejects those hypotheses that have a p-value ≤ 0.004761904761904762.
Holm's procedure rejects those hypotheses that have a p-value ≤ 0.1.
Shaffer's procedure rejects those hypotheses that have a p-value ≤ 0.004761904761904762.
Bergmann's procedure rejects these hypotheses:

- LWRK vs. MS

- LWRK vs. SPT

- LWRK vs. WINQ

- LWRK vs. —(3) AS—model—AS-model-0

- LWRK vs. —(0) NN—Baseline-model-0

- LWRK vs. —(3) AS—model—AS more rules-model-1

- MS vs. SPT

- MS vs. WINQ

- MS vs. —(3) AS—model—AS-model-0

- MS vs. —(0) NN—Baseline-model-0

- MS vs. —(3) AS—model—AS more rules-model-1

- SPT vs. WINQ

- SPT vs. —(3) AS—model—AS-model-0

- SPT vs. —(0) NN—Baseline-model-0

- SPT vs. —(3) AS—model—AS more rules-model-1

- WINQ vs. —(3) AS—model—AS-model-0

- WINQ vs. —(0) NN—Baseline-model-0

- WINQ vs. —(3) AS—model—AS more rules-model-1

- —(3) AS—model—AS-model-0 vs. —(0) NN—Baseline-model-0

- —(0) NN—Baseline-model-0 vs. —(3) AS—model—AS more rules-model-1

Table 7: Adjusted p-values

i	hypothesis	unadjusted p	p_{Neme}	p_{Holm}	p_{Shaf}	p_{Berg}
1	SPT vs .—(0) NN—Baseline-model-0	4.4553413308992614E-133	9.35621679488845E-132	9.35621679488845E-132	9.35621679488845E-132	9.35621679488845E-132
2	—(0) NN—Baseline-model-0 vs .—(3) AS—model—AS more rules-model-1	1.9748327423595857E-106	4.14714875895513E-105	3.949665484719172E-105	2.9622491135393784E-105	2.9622491135393784E-105
3	—(3) AS—model—AS-model-0 vs .—(0) NN—Baseline-model-0	1.3627983531605862E-102	2.861876541637231E-101	2.589316871005114E-101	2.0441975297408793E-101	1.499078188476645E-101
4	SPT vs .WINQ	3.471524840767506E-77	7.290202165611763E-76	6.248744713381511E-76	5.207287261151287E-76	5.207287261151287E-76
5	LWRK vs .—(0) NN—Baseline-model-0	3.294354664957412E-68	6.918144796410565E-67	5.600402930427606E-67	4.941531997436118E-67	2.964919198461671E-67
6	MS vs .SPT	1.0789706243090323E-61	2.265838311048968E-60	1.7263529988944517E-60	1.618455936635485E-60	1.1868676867399356E-60
7	WINQ vs .—(3) AS—model—AS more rules-model-1	2.4448828830165487E-57	5.1342540543347525E-56	3.667324324524823E-56	3.667324324524823E-56	2.444882883016549E-56
8	WINQ vs .—(3) AS—model—AS-model-0	1.512608531627457E-54	3.176477916411661E-53	2.117651944278440E-53	1.6638693847902032E-53	1.0588259721392202E-53
9	MS vs .—(3) AS—model—AS more rules-model-1	3.751570720872638E-44	7.87829851383254E-43	4.877041937134429E-43	4.126727792959015E-43	2.6260995046108465E-43
10	MS vs .—(3) AS—model—AS-model-0	1.024394027605593E-41	2.1512274579717455E-40	1.2292728331267116E-40	1.1268334303661524E-40	5.121970138027965E-41
11	LWRK vs .WINQ	1.3257563434190097E-30	2.78408832117992E-29	1.4583319777609106E-29	1.4583319777609106E-29	7.954538060514058E-30
12	LWRK vs .MS	2.6054990778615172E-21	5.471548063509187E-20	2.6054990778615174E-20	2.6054990778615174E-20	1.3027495389307587E-20
13	MS vs .—(0) NN—Baseline-model-0	1.5231867349708248E-15	3.198692143438732E-14	1.3708680614737424E-14	1.3708680614737424E-14	1.3708680614737424E-14
14	LWRK vs .SPT	1.281569344021574E-12	2.6912956224453053E-11	1.0252554752172591E-11	8.970985408151018E-12	8.970985408151018E-12
15	WINQ vs .—(0) NN—Baseline-model-0	2.6336714755298126E-9	5.5307100986126066E-8	1.843570032870869E-8	1.843570032870869E-8	1.053468590211925E-8
16	LWRK vs .—(3) AS—model—AS more rules-model-1	8.196700267193257E-6	1.721307056110584E-4	4.918020160315954E-5	4.918020160315954E-5	3.278680106877303E-5
17	LWRK vs .—(3) AS—model—AS-model-0	5.042055036357275E-5	0.0010588315576350279	2.5210275181786375E-4	2.5210275181786375E-4	1.5126165109071826E-4
18	SPT vs .—(3) AS—model—AS-model-0	0.0023453295675642475	0.049251920918849196	0.00938131827025699	0.00938131827025699	0.00938131827025699
19	SPT vs .—(3) AS—model—AS more rules-model-1	0.008381504991182207	0.17601160481482636	0.02514451497354662	0.02514451497354662	0.016763009982364414
20	MS vs .WINQ	0.04316499633141476	0.9064649229597099	0.08632999266282952	0.08632999266282952	0.08632999266282952
21	—(3) AS—model—AS-model-0 vs .—(3) AS—model—AS more rules-model-1	0.68451351005511373	14.374783711157884	0.68451351005511373	0.68451351005511373	0.68451351005511373