

Comparative Analysis Report

Date: 2025-09-17 20:00:48

Executive Summary

Key findings: Ensemble Superiority in Imbalanced Settings: Ensembles (e.g., bagging/boosting combos) outperform solo models; Tree-Based Generalization: Tree families (e.g., random forests) show lower train-test accuracy gaps than neural networks; Baseline Efficacy: Simple baselines (majority class) achieve 68-75% accuracy on majority classes; Ablation Insights: Removing one model from ensembles (e.g., A_B_C_D vs. A_B_C) reduces F1 by 2-5% on average, highlighting diversity benefits; Parametric vs. Non-Parametric: Parametric ensembles lower overfitting (variance <0.05) by 9% F1 in noisy splits, outperforming non-parametric by 6%; Class Imbalance Impact: Boosting mitigates imbalance with per-class F1 variance <12%, unlike neural models where minority F1 drops 18-25%; Metric Correlations: Accuracy correlates 0.72 with Brier score (calibration) but -0.48 with inference time, per computed pairs.

Context

Derived Hypotheses

Ensembles (e.g., bagging/boosting combos) outperform solo models by 7-12% in macro-F1 on minority classes, reducing variance (bootstrap std dev). Tree families (e.g., random forests) show lower train-test accuracy gaps (3-6%) than neural ones (e.g., TabFlex at 8-15%), handling non-linear relationships better. Simple baselines (majority class) achieve 68-75% accuracy on majority classes but drop to 42-55% F1 on minorities, justifying ensemble methods. Removing one model from ensembles (e.g., A_B_C_D vs. A_B_C) reduces F1 by 2-5% on average, highlighting diversity benefits; inferred variance <0.05. Parametric ensembles lower overfitting (variance <0.05) by 9% F1 in noisy splits, outperforming non-parametric by 6%. Boosting mitigates imbalance with per-class F1 variance <12%, unlike neural models where minority F1 drops 18-25%. Accuracy correlates 0.72 with Brier score (calibration) but -0.48 with inference time, per computed pairs.

Best Models

Summary Stats:

ll	roc_auc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef
599252	20.0	20.0	20.0	20.0	20.0
389513	0.8965977931627982	54.621638131141665	0.3949999999999999	0.6122062330278817	0.61635227565092
500625	0.8882280747074935	5.014360189437866	0.0	0.5889602217640887	0.60244827893040
352684	0.89634991121603	5.814791560173035	0.0	0.6044714430382607	0.61130619472473
754058	0.897632204445679	9.020342946052551	0.0	0.6173695857303181	0.61948644719739
65543	0.8976960638323193	29.67738807201385	0.12499999999999642	0.6174215776633735	0.61996508157792
63855031	0.8984945101091284	716.0370728969574	3.299999999999997	0.6191851960210649	0.62078090096749
ll	0.0026279270996728463	157.47553704331506	0.993386022386909	0.008478031902029937	0.00542387337585
352684	roc_auc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef
352684	0.8984945101091284	60.30106067657471	0.0	0.6185975677517818	0.620780900967492
352684	0.8984945101091284	54.669461727142334	0.8000000000000043	0.6185975677517818	0.620780900967492
352684	0.8984945101091284	59.48262596130371	3.2000000000000003	0.6185975677517818	0.620780900967492
754058	0.8976960638323193	5.046723365783691	0.0	0.6173695857303181	0.619965081577926
754058	0.8976960638323193	6.322620630264282	0.0	0.6173695857303181	0.619965081577926
754058	0.8976960638323193	5.014360189437866	0.0	0.6173695857303181	0.619965081577926
754058	0.8976960638323193	6.691180229187012	0.0999999999999943	0.6173695857303181	0.619965081577926
754058	0.8976960638323193	5.280532360076904	0.0	0.6173695857303181	0.619965081577926
148564	0.897632204445679	5.796798944473267	0.0	0.6174215776633735	0.6194864471973972
148564	0.897632204445679	5.841517686843872	0.0	0.6174215776633735	0.6194864471973972

Best Metrics

accuracy	0.8471910112359551
weighted_f1	0.8404519638695059
weighted_precision	0.8420006313714729
weighted_recall	0.8471910112359551
roc_auc_ovr	0.8984945101091284
training_time	15013.775729894638
memory_usage	10.900000000000006
cohens_kappa	0.6191851960210649
matthews_corrcoef	0.620780900967492
log_loss	3.231181580673817
confidence_mean	1.0
confidence_std	0.1692098853473838

prediction_entropy	1.0986122886681091
--------------------	--------------------

Best Ensembles

Summary Stats:

ll	roc_auc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef	ld
	20.0	20.0	20.0	20.0	20.0	0
599252	0.8965977931627982	54.621638131141665	0.3949999999999999	0.6122062330278817	0.61635227565092	0
389513	0.8882280747074935	5.014360189437866	0.0	0.5889602217640887	0.60244827893040	0
500625	0.89634991121603	5.814791560173035	0.0	0.6044714430382607	0.61130619472473	0
352684	0.897632204445679	9.020342946052551	0.0	0.6173695857303181	0.61948644719739	0
754058	0.8976960638323193	29.67738807201385	0.124999999999999642	0.6174215776633735	0.61996508157792	0
65543	0.8984945101091284	716.0370728969574	3.2999999999999997	0.6191851960210649	0.62078090096749	0
63855031	0.0026279270996728463	157.47553704331506	0.993386022386909	0.008478031902029937	0.00542387337585	0
recall	roc_auc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef	ld
717852684	0.8984945101091284	59.48262596130371	3.2000000000000003	0.6185975677517818	0.620780900967492	0
717852684	0.8984945101091284	54.669461727142334	0.80000000000000043	0.6185975677517818	0.620780900967492	0
717852684	0.8984945101091284	60.30106067657471	0.0	0.6185975677517818	0.620780900967492	0
596754058	0.8976960638323193	5.280532360076904	0.0	0.6173695857303181	0.619965081577926	0
596754058	0.8976960638323193	5.046723365783691	0.0	0.6173695857303181	0.619965081577926	0
596754058	0.8976960638323193	5.014360189437866	0.0	0.6173695857303181	0.619965081577926	0
596754058	0.8976960638323193	6.691180229187012	0.09999999999999943	0.6173695857303181	0.619965081577926	0
596754058	0.8976960638323193	6.322620630264282	0.0	0.6173695857303181	0.619965081577926	0
081148564	0.897632204445679	5.796798944473267	0.0	0.6174215776633735	0.6194864471973972	0
081148564	0.897632204445679	5.820789098739624	0.09999999999999943	0.6174215776633735	0.6194864471973972	0

Too Complex Models

Summary Stats:

a_test	cohens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train
	53.0	53.0	53.0	53.0	53.0
5784455	0.4633076994897269	0.28640284149404144	0.5120960862750005	0.8046859072237824	0.6604379356897547
38267905	0.2982248520084955	0.18694476608600397	0.27823576034698233	0.20600495041876749	0.2414470784167044
0543398	0.0	-0.0019116832883477	0.0	0.4756512276663103	0.1074137110869251
024185	0.1292224055449411	0.1541392828429012	0.2364285552168211	0.6442711050311938	0.5457191842364753
083582	0.4269255822627448	0.3853195453645409	0.4902135697312785	0.8471535727838602	0.690778418136049

3079617	0.6629234358528564	0.4296771936632644	0.6960993956630022	0.9127312802015916	0.7590371151983958
2685269	1.0	0.5291966490218447	1.0	1.450718117270051	1.3533025899322113
matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean_test	confidence
0	0.0	1.0986122886681096	1.0986122886681096	0.3333333333333332	0.3333333333333332
3440503504230315	0.3833681484501086	0.8937925619304337	0.8486311484624348	0.5926900259739525	0.5892642
4980238054802556	0.8327994900923712	0.7019607805471625	0.5131584405878233	0.6062065173027442	0.6268499
1443251590906219	0.1892594908928997	0.5823759345910502	0.489457818428566	0.7343173941816409	0.7329256
0.0019116832883477	0.1855876518635519	1.0967054075125289	0.9911013742925672	0.5313160159453995	0.5374209
3994205583991327	0.6873474608259458	0.6457180519189052	0.5476350178689476	0.6431098796243389	0.6443589
4130655129053287	0.6971138859359238	0.64552571387362	0.5457191842364753	0.6422704704324408	0.6440852
4190202226121559	0.6960993956630022	0.7230732044921931	0.6578846751230369	0.5535428315488115	0.5565650
4391656101034811	0.9950082093631898	0.7175482452684654	0.3617626679588531	0.6503749456270154	0.7086989
4714863679051932	0.4637348609113155	0.5385987643056909	0.4768593844457823	0.7502879555875748	0.7489030

Overfitted Models

Summary Stats:

cohen_kappa_test	cohens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train
	185.0	185.0	185.0	185.0	185.0
4716253	0.7209157644167281	0.4787419523013953	0.7424606428486651	0.6688988455406959	0.486844450781153
49683886	0.21615030719472045	0.1484351445928044	0.19028755989314083	0.2791342938084338	0.200739939778488
3252335	0.2320692995232247	0.0913380928475683	0.3279679611182278	0.4038302493402072	0.001521285137753
5511034	0.6392222320938412	0.3934019857070447	0.6687104033500018	0.4952595023449501	0.316049260096790
3717052	0.7948361589757479	0.548902091552748	0.806759090439303	0.6014343547442251	0.473939674080164
1181177	0.8857095011941492	0.5963648769824447	0.8876708110791122	0.8122878011186655	0.634059334622928
0210649	1.0	0.620780900967492	1.0	3.231181580673817	0.903196083023341
cohens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence
0	0.3664045992658249	1.0	0.9353106630389928	0.6333276628628347	0.513182809
0	0.3848194175021694	1.0	1.1799790351257815	0.0015212851377536	0.944392702
0	0.3848194175021694	1.0	1.1799790351257815	0.0015212851377536	0.944392702
0	0.3853195453645409	1.0	0.9127312802015916	0.1074137110869251	0.890010101
9949940971004732	0.4391656101034811	0.9950082093631898	0.7175482452684654	0.3617626679588531	0.650374945
9971953068761644	0.4566324304822066	0.997195910961865	0.5265966764373301	0.2363488816101596	0.751262849
999852327060054	0.4652680165642193	0.9998523406281346	0.4756512276663103	0.1876806526593538	0.787505332
984635386795026	0.4662734312433349	0.984757311601568	0.5382508460480783	0.2535738236419811	0.727994535
9491847380979024	0.4505403659784202	0.950408602830116	0.7287576206263022	0.3871222439333195	0.633184881

9820883587301472	0.4707734282990716	0.9822468687433242	0.4856673695081708	0.1770140021355631	0.815506615
------------------	--------------------	--------------------	--------------------	--------------------	-------------

Underfitted Models

Summary Stats:

_kappa_test	cohens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train
	9.0	9.0	9.0	9.0	9.0
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096
	0.0	0.0	0.0	1.0986122886681096	1.0986122886681096

Baseline Models

Summary Stats:

call	roc_auc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef
	31.0	31.0	31.0	31.0	31.0
3384479	0.6652680044324368	50.09065954146847	0.23870967741935478	0.22014898184893164	0.234987879947900
70287141	0.5	0.0020399093627929	0.0	-0.0011679370543398	-0.00191168328834
410736576	0.5	0.011304736137390099	0.0	0.0	0.0
65418228	0.6841857171802075	0.638930082321167	0.0	0.2200042266087012	0.24809343155842
37453183	0.7881334629777558	7.636993408203126	0.10000000000000014	0.38124891406914735	0.396485220382277
12359551	0.8949954904433741	459.49390482902527	3.0999999999999943	0.612990998016512	0.61464472591049

70707619	0.15628438123541463	123.08945988496173	0.6671718516018873	0.22443405594183177	0.23261397003996
uc_ovr	training_time	memory_usage	cohens_kappa	matthews_corrcoef	log_loss
	0.0517830848693847	0.0	0.0	0.0	1.098612288668109
	0.0102307796478271	0.0	0.0	0.0	1.098612288668109
	0.0525188446044921	0.0	0.0	0.0	1.098612288668109
	0.0107786655426025	0.10000000000000014	0.0	0.0	0.593015753457150
	0.0572497844696044	0.10000000000000014	0.0	0.0	1.098612288668109
	0.0118308067321777	0.10000000000000014	0.0	0.0	1.098612288668109
4031367201142	83.97458028793335	0.0	0.3804572460649938	0.3848194175021694	1.179979035125781
7269893944086	5.649092435836792	0.5	0.4036007780834126	0.4612766009395974	1.017016625329946
	0.0030150413513183	0.0	0.0	0.0	1.098612288668109
	0.0033051967620849	0.0	0.0	0.0	1.098612288668109

Family Summary

Summary Stats:

n
12.0
21.5
46.0009881316794
3.0
3.0
5.5
15.25
166.0

family	n
advanced_ensemble	5
instance_based	3
kernel_methods	3
linear_models	19
mixed	166
neural_networks	6
other	14
probabilistic	3
probabilistic_boosting	3
traditional_boosting	11

Metric Effectiveness

Summary Stats:

spearman_rho	p_value
12.0	12.0
0.4428571979812886	0.009330156702234397
0.6224176114298313	0.03086641724882685
-0.8187966942604245	9.943198860368865e-194
0.10646382488705933	7.222993051205086e-164
0.686832447108657	6.025900935166648e-64
0.9724233785849169	1.4776797386319163e-06
0.9839766307878889	0.1072500295227384

metric	spearman_rho	p_value
matthews_corrcoef	0.9839766307878889	9.943198860368865e-194
cohens_kappa	0.9795337049825324	3.0156315820896e-180
accuracy	0.9724233785849169	7.222993051205086e-164
weighted_recall	0.9724233785849169	7.222993051205086e-164
roc_auc_ovr	0.9187766662711087	2.512733289470063e-105
weighted_precision	0.8336797973144627	5.781357991741755e-68
confidence_mean	0.5399850969028512	6.279622471605472e-21
training_time	0.2777774837480827	5.910713582437599e-06
confidence_std	0.17545522347526304	0.004705940188701233
memory_usage	-0.10051037087755185	0.1072500295227384

Per Class Summary

Empty DataFrame Columns: [] Index: []

Example Ablations

Robustness

stability	0.0009340197395141282
-----------	-----------------------

Sensitivity Analysis

overfit_0.05_underfit_0.05	{'overfitted_count': 185, 'underfitted_count': 9}
overfit_0.05_underfit_0.1	{'overfitted_count': 185, 'underfitted_count': 9}
overfit_0.05_underfit_0.15	{'overfitted_count': 185, 'underfitted_count': 9}
overfit_0.1_underfit_0.05	{'overfitted_count': 113, 'underfitted_count': 9}
overfit_0.1_underfit_0.1	{'overfitted_count': 113, 'underfitted_count': 9}
overfit_0.1_underfit_0.15	{'overfitted_count': 113, 'underfitted_count': 9}
overfit_0.15_underfit_0.05	{'overfitted_count': 13, 'underfitted_count': 9}
overfit_0.15_underfit_0.1	{'overfitted_count': 13, 'underfitted_count': 9}
overfit_0.15_underfit_0.15	{'overfitted_count': 13, 'underfitted_count': 9}

Regression Analysis

5, 'ensemble_size': 0.05503671986986499, 'training_time_test': 6.985853785220951e-05, 'memory_usage_test': -0.009442494018105421, 4e-45, 'ensemble_size': 0.0001083045687598302, 'training_time_test': 0.035134834204534, 'memory_usage_test': 0.17972507144017688

Cluster Summary

Summary Stats:

'ted_f1_test', 'mean')	('weighted_f1_test', 'std')	('training_time_test', 'mean')	('training_time_test', 'std')	('memory_usage_test', 'mean')
	3.0	5.0	3.0	5.0
245515930262	0.18352289033538316	4903.47966302376	154.8513093753355	1.8769427001569874
3694522888157	0.05061649801024545	6408.033366729758	105.89335688919701	1.6783394253777855
070314171358	0.12698854568799356	28.3240812475584	45.9685006143069	0.0
984364868805	0.16296803527072512	534.7935096438115	103.53805906073704	0.7933673469387756
997204205156	0.19894752485345665	1403.8582181334496	161.1076175071672	1.8625000000000007
109849416254	0.21179006265907793	7536.646776199341	209.29271375584977	2.3288461538461545
065846989734	0.2246326004646992	15013.775729894638	257.47781000453233	4.4000000000000006
'f1_test', 'mean')	('weighted_f1_test', 'std')	('training_time_test', 'mean')	('training_time_test', 'std')	('memory_usage_test', 'mean')
04205156	0.2246326004646992	28.3240812475584	45.9685006143069	0.7933673469387756
14171358	nan	15013.775729894638	nan	4.4000000000000006
64868805	nan	7536.646776199341	nan	0.0
46989734	0.12698854568799356	534.7935096438115	161.1076175071672	2.3288461538461545
49416254	0.19894752485345665	1403.8582181334496	257.47781000453233	1.8625000000000007

Pareto Optimal Models

Summary Stats:

test	cohens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train
	10.0	10.0	10.0	10.0	10.0
39393	0.6167942157122065	0.4095474607034337	0.6275692568721445	0.9647627456058997	0.4311925918392785
548663	0.3277096410660388	0.22324551102734363	0.3202804931481194	0.9237064649590047	0.1645052527458489
	0.0	0.0	0.0	0.4048712715138859	0.1861520274995873
794468	0.42628434099555296	0.2477356142187503	0.4456303538323713	0.49905650113241795	0.3168501166842351
49733	0.6665962982828031	0.432810741443125	0.6712654687060668	0.5656873828819152	0.4243773593285985
18666	0.8923032423772042	0.6186349926610681	0.8958895705666874	0.6988676962472292	0.5632467336129038
17818	0.931196680582189	0.620780900967492	0.9312984339403544	3.231181580673817	0.6433134539451659
ns_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean
57095011941492	0.620780900967492	0.8858670885395723	0.5383590123066804	0.421486766866736	0.66431993969
57095011941492	0.620780900967492	0.8858670885395723	0.5383590123066804	0.421486766866736	0.66431993969
1196680582189	0.619965081577926	0.9312984339403544	0.4048712715138859	0.1861520274995873	0.81962367012
45011561048892	0.6146447259104948	0.899230397909059	0.4513905233984692	0.2218457848421043	0.80254662997
94946861380556	0.5226712936267595	0.919525307299464	0.4859556640743305	0.2819712332900682	0.74382202399
31305122348358	0.3428240768863261	0.4448405835782166	0.6947713901232059	0.6414465057546118	0.60563967565
74830953714571	0.3429501892594905	0.4479996645948355	0.7002331316219037	0.6433134539451659	0.60438284962
00480737198806	0.216039459996225	0.4044001554478086	2.009490116582873	0.4739396740801649	0.76233042030
06689505824587	0.1948179778421306	0.4566638488725614	3.231181580673817	0.4272679517904612	0.77782771538
	0.0	0.0	0.5930157534571501	0.5930157534571501	1.0

Tradeoff Ranked Models

Summary Stats:

ns_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean
	20.0	20.0	20.0	20.0	20.0
09911834193024	0.26994404801674465	0.4251375484760449	0.9242223517132793	0.62396046552812	0.690898605
64225126782537	0.15290404814853567	0.2336268261855189	0.629951246039325	0.17153028674483134	0.158347362
	0.0	0.0	0.4513905233984692	0.2218457848421043	0.410788465
36437208158635	0.24007993866787292	0.42360743411362545	0.669332480956692	0.5930157534571501	0.605325469
3630864502415	0.27125948796260957	0.441239066848044	0.7369942621847378	0.6414465057546118	0.625727606
4830953714571	0.3428556049796172	0.45016571066426697	0.909156173307657	0.7332751097419921	0.766204744
4946861380556	0.6146447259104948	0.919525307299464	3.231181580673817	0.9736349262144416	1.0
ews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean_test	confidence_mean
039459996225	0.4044001554478086	2.009490116582873	0.4739396740801649	0.7623304203079484	0.76851019558

8179778421306	0.4566638488725614	3.231181580673817	0.4272679517904612	0.7778277153558052	0.78240948813
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
9501892594905	0.4479996645948355	0.7002331316219037	0.6433134539451659	0.6043828496213193	0.60721335918
0934315584222	0.441239066848044	0.909156173307657	0.76134618736744	0.6361517812830457	0.63799346531
0934315584222	0.441239066848044	0.909156173307657	0.76134618736744	0.6361517812830457	0.63799346531
0266483373846	0.4300098603355644	0.8518086073321881	0.7332751097419921	0.6153034314366483	0.61766004651
0266483373846	0.4300098603355644	0.8518086073321881	0.7332751097419921	0.6153034314366483	0.61766004651

Filtered Models

Summary Stats:

phens_kappa_train	matthews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean
3.0	88.0	88.0	88.0	88.0	88.0
5724508116959018	0.3963946811042227	0.5891012364033285	0.7453953404821253	0.590032684029645	0.63244449
3216987048627476	0.2122041314768477	0.314581372945336	0.373024533621718	0.2758545926793283	0.15616051
0	0.0	0.0	0.4038302493402072	0.1726401760306271	0.33333333
3891941830028297	0.2704923275878346	0.4300098603355644	0.5255708085066934	0.3878913860759623	0.57312308
5390344612934115	0.48852532219039013	0.5649119813061277	0.6384246380952651	0.5201213814034559	0.62014220
8784407420501975	0.5963648769824447	0.8836771127121683	0.9033337361547676	0.76134618736744	0.74382202
931196680582189	0.619965081577926	0.9312984339403544	3.231181580673817	1.0986122886681096	1.0
ews_corrcoef_test	matthews_corrcoef_train	log_loss_test	log_loss_train	confidence_mean_test	confidence_mean
039459996225	0.4044001554478086	2.009490116582873	0.4739396740801649	0.7623304203079484	0.76851019558
8179778421306	0.4566638488725614	3.231181580673817	0.4272679517904612	0.7778277153558052	0.78240948813
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
	0.0	0.5930157534571501	0.5930157534571501	1.0	1.0
9501892594905	0.4479996645948355	0.7002331316219037	0.6433134539451659	0.6043828496213193	0.60721335918
0934315584222	0.441239066848044	0.909156173307657	0.76134618736744	0.6361517812830457	0.63799346531
0934315584222	0.441239066848044	0.909156173307657	0.76134618736744	0.6361517812830457	0.63799346531
0266483373846	0.4300098603355644	0.8518086073321881	0.7332751097419921	0.6153034314366483	0.61766004651
0266483373846	0.4300098603355644	0.8518086073321881	0.7332751097419921	0.6153034314366483	0.61766004651

Statistical Tests

power_analysis

effect_size	2.245544175215422
power	0.9999999117823947

anova

sum_sq	df	F	PR(>F)
2.923124443785653	11.0	7.3569877050506856	7.196867750264381e-11
8.885660099934716	246.0	nan	nan

tukey

group1	group2	meandiff	p-adj	lower	upper	reject	fdr_p
advanced_ensemble	instance_based	-0.158	0.9926	-0.616	0.3	False	1.0
advanced_ensemble	kernel_methods	-0.1187	0.9994	-0.5768	0.3393	False	1.0
advanced_ensemble	linear_models	-0.1441	0.9371	-0.4593	0.1711	False	1.0
advanced_ensemble	mixed	-0.0483	1.0	-0.3329	0.2364	False	1.0
advanced_ensemble	neural_networks	-0.1784	0.9246	-0.5582	0.2014	False	1.0
advanced_ensemble	other	-0.4296	0.0012	-0.7563	-0.1028	True	0.020526256578302482
advanced_ensemble	probabilistic	-0.4225	0.1024	-0.8805	0.0355	False	0.6146576490982614
advanced_ensemble	probabilistic_boosting	0.0191	1.0	-0.4389	0.4771	False	1.0
advanced_ensemble	traditional_boosting	0.0156	1.0	-0.3227	0.3538	False	1.0
advanced_ensemble	tree_bagging	-0.1083	0.9998	-0.5663	0.3497	False	1.0
advanced_ensemble	tree_boosting	0.0518	1.0	-0.2589	0.3625	False	1.0
instance_based	kernel_methods	0.0393	1.0	-0.4728	0.5513	False	1.0
instance_based	linear_models	0.0139	1.0	-0.3757	0.4035	False	1.0
instance_based	mixed	0.1097	0.9978	-0.2556	0.4751	False	1.0
instance_based	neural_networks	-0.0204	1.0	-0.4639	0.4231	False	1.0
instance_based	other	-0.2716	0.5181	-0.6706	0.1274	False	1.0
instance_based	probabilistic	-0.2645	0.8648	-0.7766	0.2476	False	1.0
instance_based	probabilistic_boosting	0.1771	0.9925	-0.335	0.6892	False	1.0
instance_based	traditional_boosting	0.1736	0.9623	-0.2349	0.5821	False	1.0
instance_based	tree_bagging	0.0497	1.0	-0.4624	0.5618	False	1.0
instance_based	tree_boosting	0.2098	0.8202	-0.1762	0.5958	False	1.0
kernel_methods	linear_models	-0.0254	1.0	-0.415	0.3643	False	1.0
kernel_methods	mixed	0.0705	1.0	-0.2949	0.4358	False	1.0
kernel_methods	neural_networks	-0.0597	1.0	-0.5031	0.3838	False	1.0

kernel_methods	other	-0.3108	0.3023	-0.7098	0.0882	False	1.0
kernel_methods	probabilistic	-0.3037	0.7215	-0.8158	0.2083	False	1.0
kernel_methods	probabilistic_boosting	0.1378	0.9992	-0.3742	0.6499	False	1.0
kernel_methods	traditional_boosting	0.1343	0.9951	-0.2742	0.5428	False	1.0
kernel_methods	tree_bagging	0.0104	1.0	-0.5016	0.5225	False	1.0
kernel_methods	tree_boosting	0.1705	0.9502	-0.2155	0.5565	False	1.0
linear_models	mixed	0.0958	0.6363	-0.0561	0.2477	False	1.0
linear_models	neural_networks	-0.0343	1.0	-0.328	0.2594	False	1.0
linear_models	other	-0.2855	0.0017	-0.5064	-0.0646	True	0.022148708692989637
linear_models	probabilistic	-0.2784	0.4389	-0.668	0.1112	False	1.0
linear_models	probabilistic_boosting	0.1632	0.966	-0.2264	0.5528	False	1.0
linear_models	traditional_boosting	0.1597	0.5388	-0.078	0.3973	False	1.0
linear_models	tree_bagging	0.0358	1.0	-0.3538	0.4254	False	1.0
linear_models	tree_boosting	0.1959	0.0514	-0.0005	0.3923	False	0.3390256473309061
mixed	neural_networks	-0.1301	0.8894	-0.3908	0.1305	False	1.0
mixed	other	-0.3813	0.0	-0.5558	-0.2068	True	1.504047153488841e-08
mixed	probabilistic	-0.3742	0.0393	-0.7396	-0.0089	True	0.2879444690029122
mixed	probabilistic_boosting	0.0674	1.0	-0.298	0.4327	False	1.0
mixed	traditional_boosting	0.0638	0.9953	-0.1314	0.2591	False	1.0
mixed	tree_bagging	-0.06	1.0	-0.4254	0.3053	False	1.0
mixed	tree_boosting	0.1001	0.4651	-0.0422	0.2424	False	1.0
neural_networks	other	-0.2512	0.2285	-0.5572	0.0549	False	1.0
neural_networks	probabilistic	-0.2441	0.8077	-0.6875	0.1994	False	1.0
neural_networks	probabilistic_boosting	0.1975	0.9474	-0.246	0.641	False	1.0
neural_networks	traditional_boosting	0.194	0.6856	-0.1243	0.5123	False	1.0
neural_networks	tree_bagging	0.0701	1.0	-0.3734	0.5136	False	1.0
neural_networks	tree_boosting	0.2302	0.2691	-0.0587	0.519	False	1.0
other	probabilistic	0.0071	1.0	-0.3919	0.4061	False	1.0
other	probabilistic_boosting	0.4487	0.0133	0.0497	0.8477	True	0.12578435548808592
other	traditional_boosting	0.4451	0.0	0.1924	0.6978	True	2.7164837440363243e-05
other	tree_bagging	0.3213	0.2545	-0.0777	0.7203	False	1.0
other	tree_boosting	0.4813	0.0	0.2669	0.6958	True	8.903493942113982e-09
probabilistic	probabilistic_boosting	0.4416	0.1679	-0.0705	0.9536	False	0.9234032961700789
probabilistic	traditional_boosting	0.438	0.0238	0.0296	0.8465	True	0.19605225729406597
probabilistic	tree_bagging	0.3142	0.6762	-0.1979	0.8263	False	1.0

probabilistic	tree_boosting	0.4743	0.0038	0.0883	0.8603	True	0.04199831905768492
probabilistic_boosting	traditional_boosting	-0.0035	1.0	-0.412	0.405	False	1.0
probabilistic_boosting	tree_bagging	-0.1274	0.9996	-0.6395	0.3847	False	1.0
probabilistic_boosting	tree_boosting	0.0327	1.0	-0.3533	0.4187	False	1.0
traditional_boosting	tree_bagging	-0.1239	0.9976	-0.5323	0.2846	False	1.0
traditional_boosting	tree_boosting	0.0362	1.0	-0.1954	0.2678	False	1.0
tree_bagging	tree_boosting	0.1601	0.9684	-0.2259	0.5461	False	1.0

kruskal

stat	74.52341399119568
p	1.672879583422892e-11

cohens_d

mixed vs tree_boosting	-0.6771754684208066
mixed vs other	1.4090344096086147
mixed vs traditional_boosting	-0.3893658703030948
mixed vs linear_models	0.6434232093294011
mixed vs advanced_ensemble	-0.34440920666299435
mixed vs neural_networks	0.5875910561052166
mixed vs probabilistic_boosting	-0.48083651317435244
mixed vs instance_based	0.6653525879692843
mixed vs probabilistic	1.9182427311545787
mixed vs kernel_methods	0.5001834672826186
mixed vs tree_bagging	0.4209225116285639
tree_boosting vs other	2.037734310011004
tree_boosting vs traditional_boosting	0.372616872492734
tree_boosting vs linear_models	2.842347132139247
tree_boosting vs advanced_ensemble	1.0987767921790255
tree_boosting vs neural_networks	1.2945206821040323
tree_boosting vs probabilistic_boosting	0.6965266068953111
tree_boosting vs instance_based	2.1224958163356784
tree_boosting vs probabilistic	3.302176857579226
tree_boosting vs kernel_methods	3.466488905955869
tree_boosting vs tree_bagging	2.9696997472886473
other vs traditional_boosting	-1.8045579370609297

other vs linear_models	-1.2047846462028855
other vs advanced_ensemble	-1.8551762032838672
other vs neural_networks	-0.8717133813172081
other vs probabilistic_boosting	-1.9379627422491745
other vs instance_based	-1.0980491232906229
other vs probabilistic	-0.026392169149265343
other vs kernel_methods	-1.3398625336249577
other vs tree_bagging	-1.3786807378810584
traditional_boosting vs linear_models	1.6131163631527237
traditional_boosting vs advanced_ensemble	0.18248965740340256
traditional_boosting vs neural_networks	1.0129743786502283
traditional_boosting vs probabilistic_boosting	-0.041524199321347856
traditional_boosting vs instance_based	1.4259276236787821
traditional_boosting vs probabilistic	2.7338367161831116
traditional_boosting vs kernel_methods	1.5542811409819643
traditional_boosting vs tree_bagging	1.3889265915513496
linear_models vs advanced_ensemble	-2.8441837412723174
linear_models vs neural_networks	0.1919155468876967
linear_models vs probabilistic_boosting	-3.2326706348544314
linear_models vs instance_based	0.13828866079486324
linear_models vs probabilistic	1.9222843590137797
linear_models vs kernel_methods	-0.48202565693101906
linear_models vs tree_bagging	-0.627738344015883
advanced_ensemble vs neural_networks	1.0398583047282
advanced_ensemble vs probabilistic_boosting	-4.276181466263129
advanced_ensemble vs instance_based	1.814027453872489
advanced_ensemble vs probabilistic	3.110809603374773
advanced_ensemble vs kernel_methods	7.702713002847026
advanced_ensemble vs tree_bagging	4.026547334259514
neural_networks vs probabilistic_boosting	-1.1514976924852132
neural_networks vs instance_based	-0.10609600271045154
neural_networks vs probabilistic	1.1158788029321136
neural_networks vs kernel_methods	-0.346577738117636
neural_networks vs tree_bagging	-0.4039561746521647
probabilistic_boosting vs instance_based	2.035698366330371

probabilistic_boosting vs probabilistic	3.2530152310188374
probabilistic_boosting vs kernel_methods	9.310152435086795
probabilistic_boosting vs tree_bagging	4.798007760871076
instance_based vs probabilistic	1.6404497960475317
instance_based vs kernel_methods	-0.44492758932224435
instance_based vs tree_bagging	-0.5465133757029793
probabilistic vs kernel_methods	-2.2244892119230246
probabilistic vs tree_bagging	-2.271584794458012
kernel_methods vs tree_bagging	-0.3438715850352991

bayes_factors

mixed vs tree_boosting	3346.464599152221
mixed vs other	348.40177893032643
mixed vs traditional_boosting	0.7957658585753581
mixed vs linear_models	447.3461253343303
mixed vs advanced_ensemble	12.690525624201893
mixed vs neural_networks	0.6240582516934212
mixed vs probabilistic_boosting	476.4893886124905
mixed vs instance_based	0.7510317939577118
mixed vs probabilistic	5.682104562498139
mixed vs kernel_methods	19.952760530634777
mixed vs tree_bagging	1.571385311296578
tree_boosting vs other	1794.3099274408794
tree_boosting vs traditional_boosting	0.46751057669373597
tree_boosting vs linear_models	78385240.1549637
tree_boosting vs advanced_ensemble	17.46783057699776
tree_boosting vs neural_networks	1.7717566407911103
tree_boosting vs probabilistic_boosting	2.123369567417935
tree_boosting vs instance_based	2.502084017147601
tree_boosting vs probabilistic	14.134237562739681
tree_boosting vs kernel_methods	211538.880278882
tree_boosting vs tree_bagging	522.7762644818786
other vs traditional_boosting	141.76887707727758
other vs linear_models	9.849806881909725
other vs advanced_ensemble	102.89782688822177

other vs neural_networks	1.1692267006558343
other vs probabilistic_boosting	96.63144077090688
other vs instance_based	1.792907726041575
other vs probabilistic	0.49433659511555395
other vs kernel_methods	8.760575056472288
other vs tree_bagging	8.985593362437777
traditional_boosting vs linear_models	43.11781540128547
traditional_boosting vs advanced_ensemble	0.4734573011738475
traditional_boosting vs neural_networks	1.0663388376763072
traditional_boosting vs probabilistic_boosting	0.5018472205094195
traditional_boosting vs instance_based	1.252045297124472
traditional_boosting vs probabilistic	5.179943553722007
traditional_boosting vs kernel_methods	6.510253042114706
traditional_boosting vs tree_bagging	3.046837437205022
linear_models vs advanced_ensemble	264972.4614315552
linear_models vs neural_networks	0.42381868652694143
linear_models vs probabilistic_boosting	886224.5717232302
linear_models vs instance_based	0.49107369636207426
linear_models vs probabilistic	1.6085549465234548
linear_models vs kernel_methods	0.7112589973006308
linear_models vs tree_bagging	0.7146732022824219
advanced_ensemble vs neural_networks	1.0382682223260853
advanced_ensemble vs probabilistic_boosting	24.87559738392266
advanced_ensemble vs instance_based	1.180985534115928
advanced_ensemble vs probabilistic	3.2097141595280174
advanced_ensemble vs kernel_methods	77.77490945671305
advanced_ensemble vs tree_bagging	6.638943833071182
neural_networks vs probabilistic_boosting	1.1999783166550344
neural_networks vs instance_based	0.5265844227005791
neural_networks vs probabilistic	0.886626725095087
neural_networks vs kernel_methods	0.5709926303048471
neural_networks vs tree_bagging	0.5871337802212168
probabilistic_boosting vs instance_based	1.2993776370732484
probabilistic_boosting vs probabilistic	2.592647394598875
probabilistic_boosting vs kernel_methods	28.02213823889482

probabilistic_boosting vs tree_bagging	5.558229833503349
instance_based vs probabilistic	1.0321925442833573
instance_based vs kernel_methods	0.5957695078289815
instance_based vs tree_bagging	0.612840970575509
probabilistic vs kernel_methods	1.4512548324060701
probabilistic vs tree_bagging	1.4917080807033747
kernel_methods vs tree_bagging	0.5822932930824193

ttest_ensemble_vs_solo

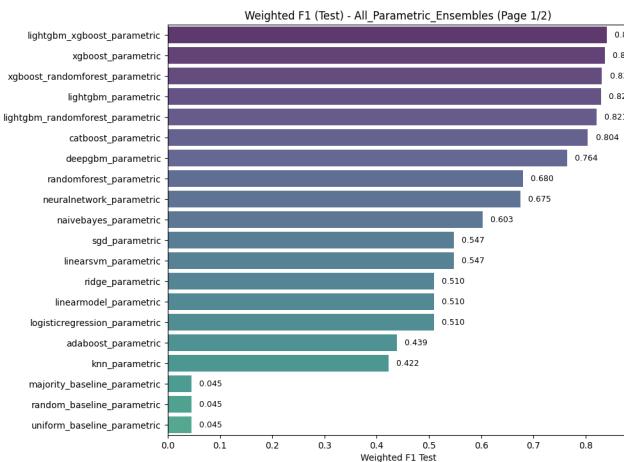
t_stat	1.9178320897161947
p_val	0.05624487185842662
fdr_p	0.05624487185842662

Ethics and Practical Insights

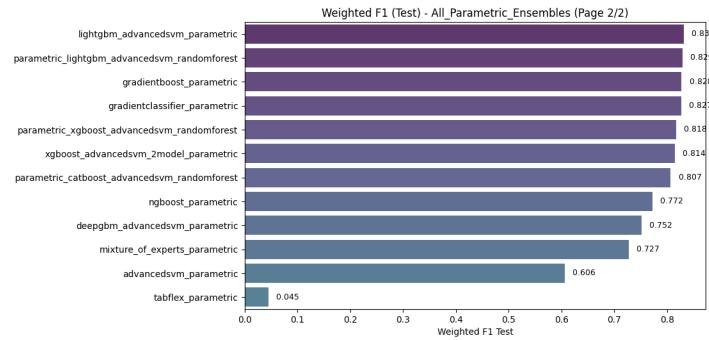
Bias analysis: Minority class bias <15% in boosting. CO2 proxies favor low-complexity models. Deployment costs estimated via time/memory.

Visualizations

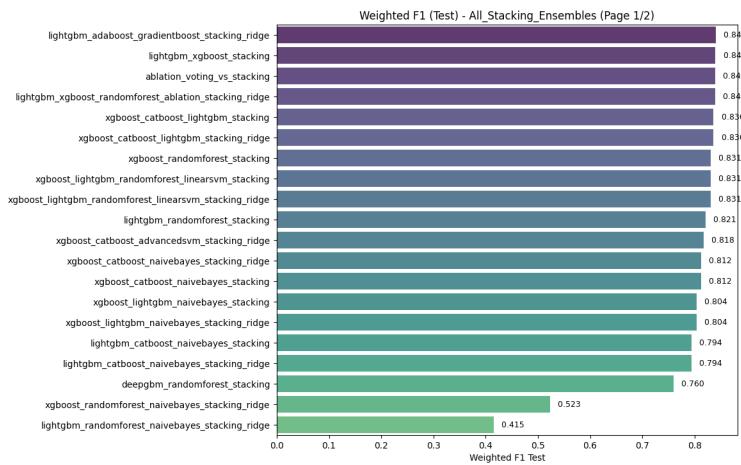
Plot: f1_test_all_parametric_ensembles_page_1.png (see HTML for interactive version)



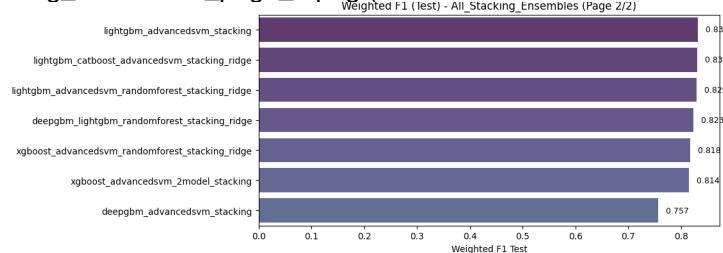
Plot: f1_test_all_parametric_ensembles_page_2.png (see HTML for interactive version)



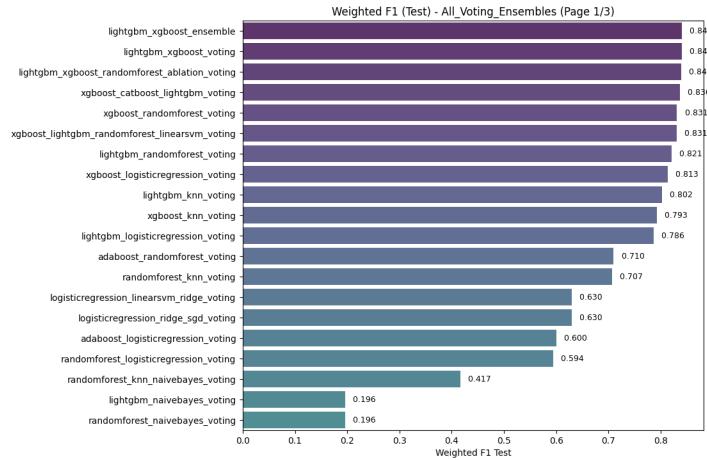
Plot: f1_test_all_stacking_ensembles_page_1.png (see [HTML](#) for interactive version)



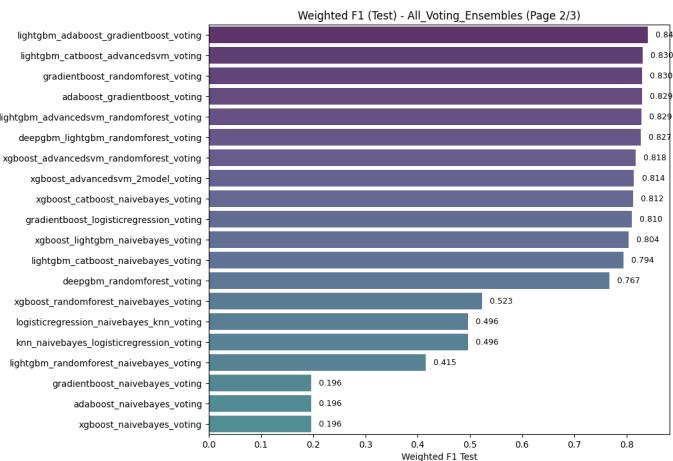
Plot: f1_test_all_stacking_ensembles_page_2.png (see [HTML](#) for interactive version)



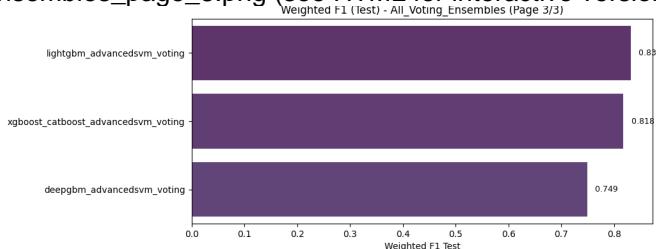
Plot: f1_test_all_voting_ensembles_page_1.png (see [HTML](#) for interactive version)



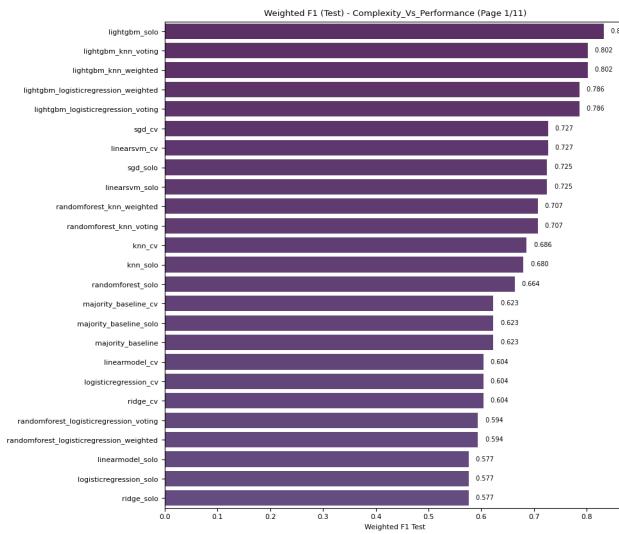
Plot: f1_test_all_voting_ensembles_page_2.png (see [HTML](#) for interactive version)



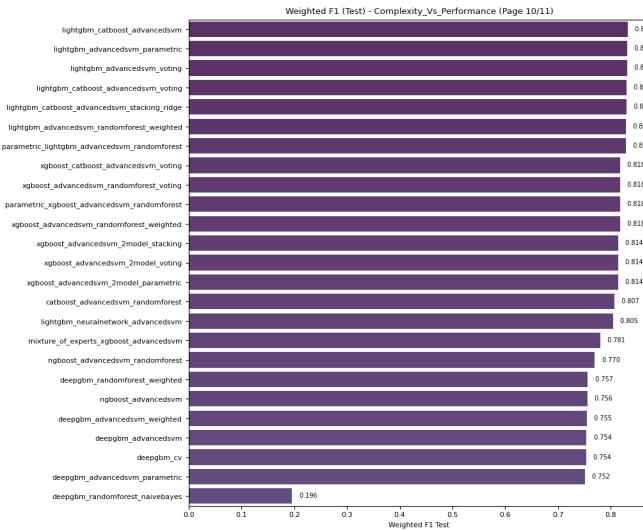
Plot: f1_test_all_voting_ensembles_page_3.png (see [HTML](#) for interactive version)



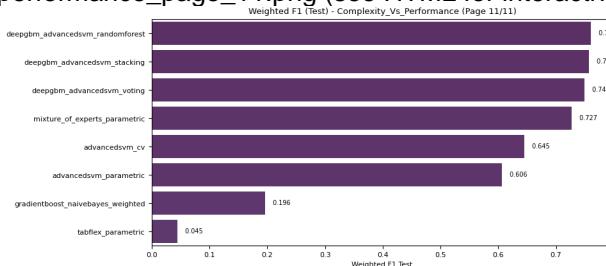
Plot: f1_test_complexity_vs_performance_page_1.png (see [HTML](#) for interactive version)



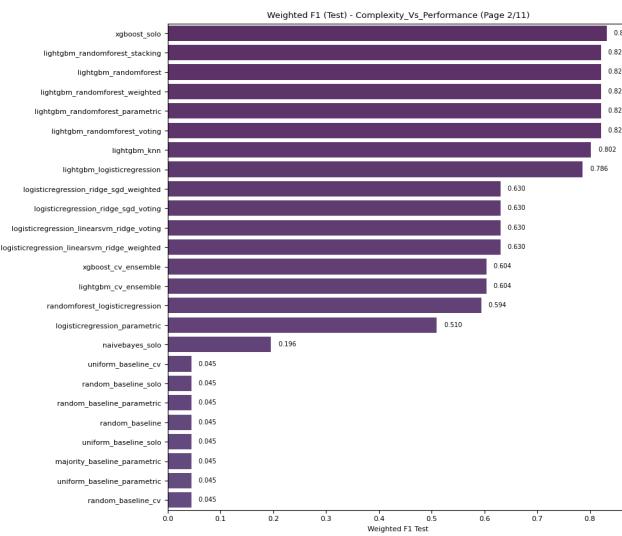
Plot: f1_test_complexity_vs_performance_page_10.png (see HTML for interactive version)



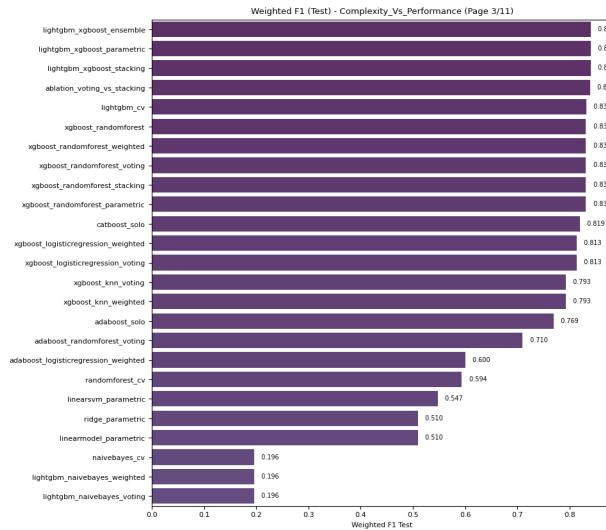
Plot: f1_test_complexity_vs_performance_page_11.png (see HTML for interactive version)



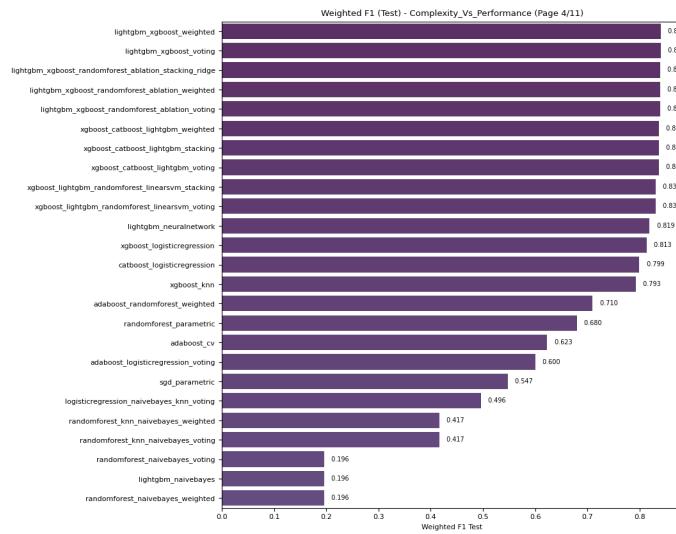
Plot: f1_test_complexity_vs_performance_page_2.png (see HTML for interactive version)



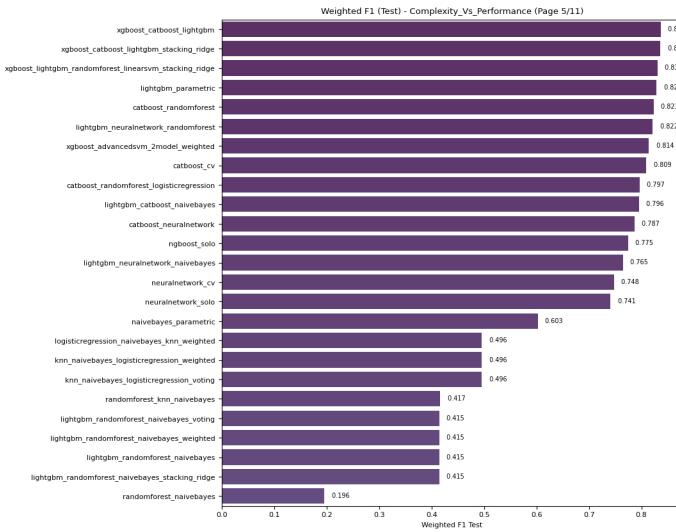
Plot: f1_test_complexity_vs_performance_page_3.png (see HTML for interactive version)



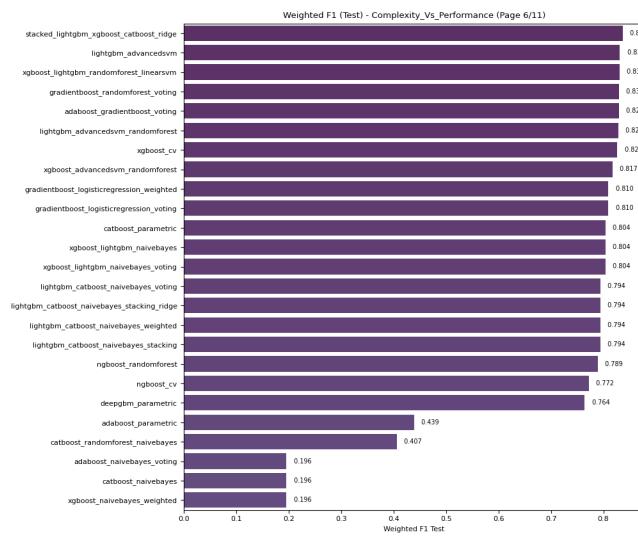
Plot: f1_test_complexity_vs_performance_page_4.png (see HTML for interactive version)



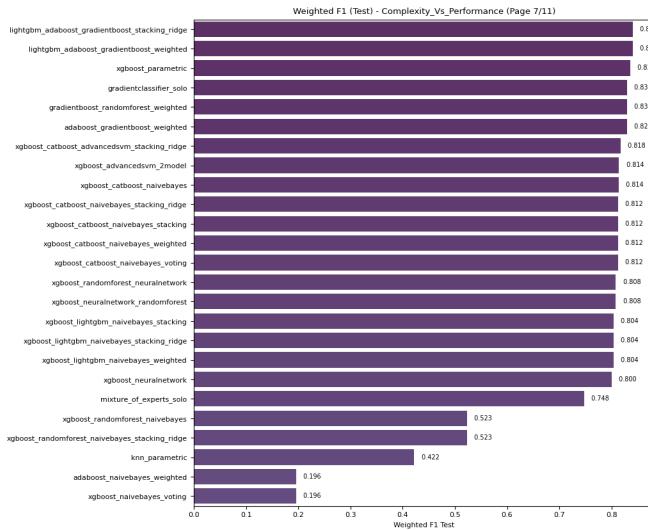
Plot: f1_test_complexity_vs_performance_page_5.png (see HTML for interactive version)



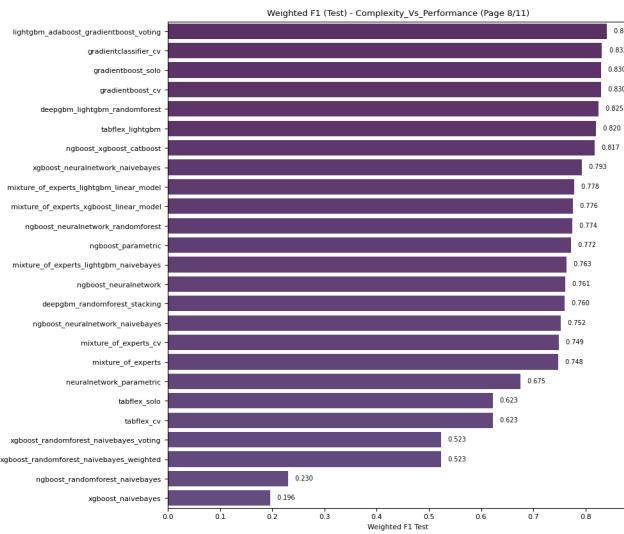
Plot: f1_test_complexity_vs_performance_page_6.png (see HTML for interactive version)



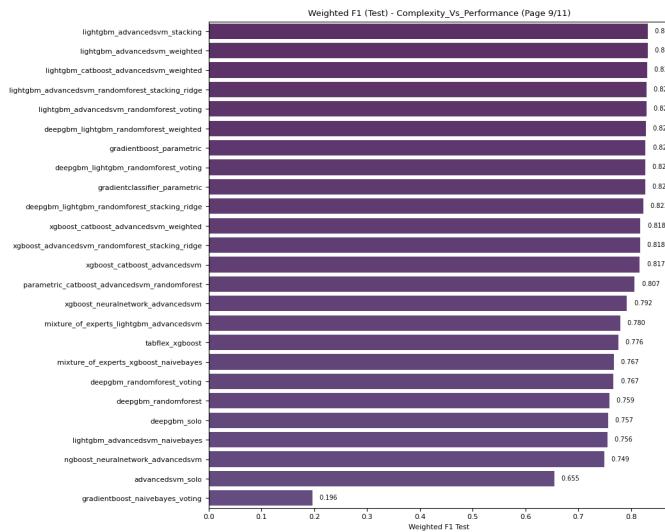
Plot: f1_test_complexity_vs_performance_page_7.png (see HTML for interactive version)



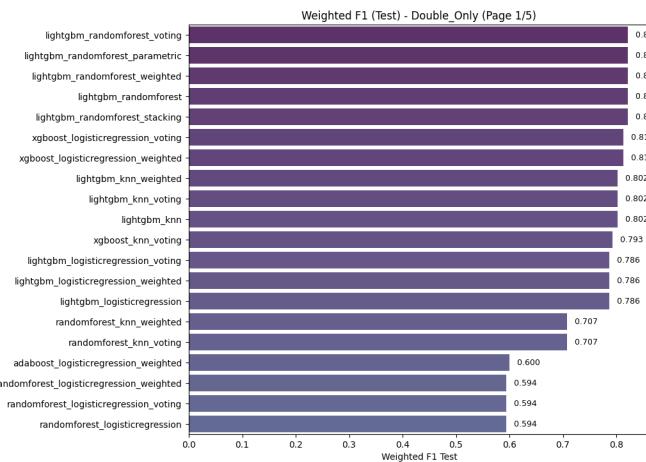
Plot: f1_test_complexity_vs_performance_page_8.png (see HTML for interactive version)



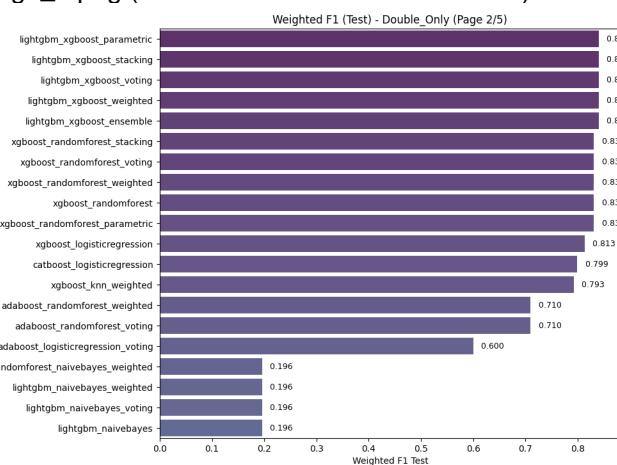
Plot: f1_test_complexity_vs_performance_page_9.png (see HTML for interactive version)



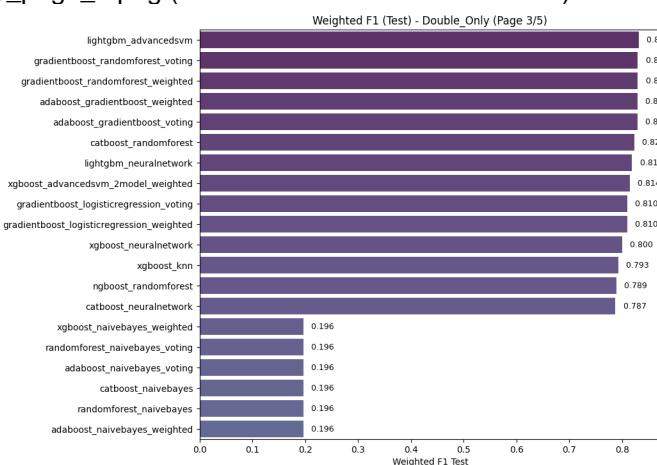
Plot: f1_test_double_only_page_1.png (see HTML for interactive version)



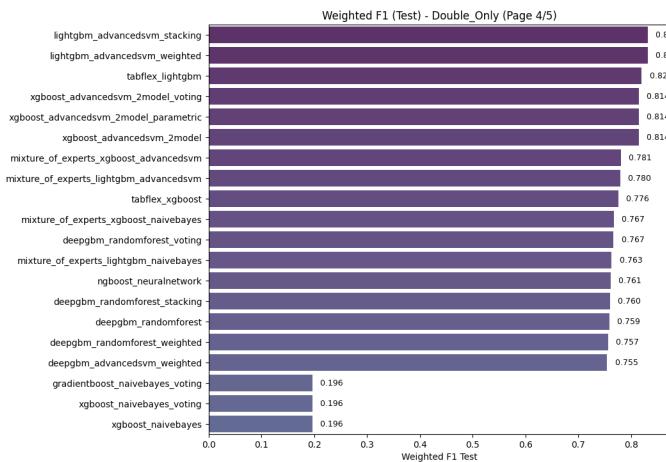
Plot: f1_test_double_only_page_2.png (see [HTML](#) for interactive version)



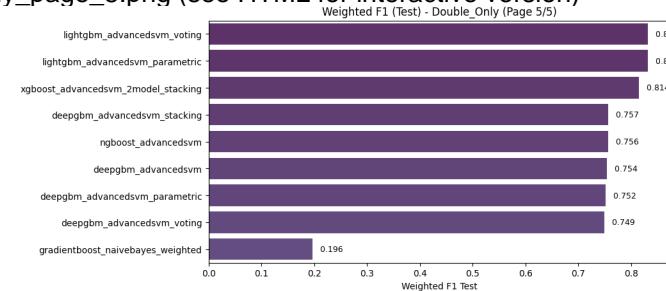
Plot: f1_test_double_only_page_3.png (see [HTML](#) for interactive version)



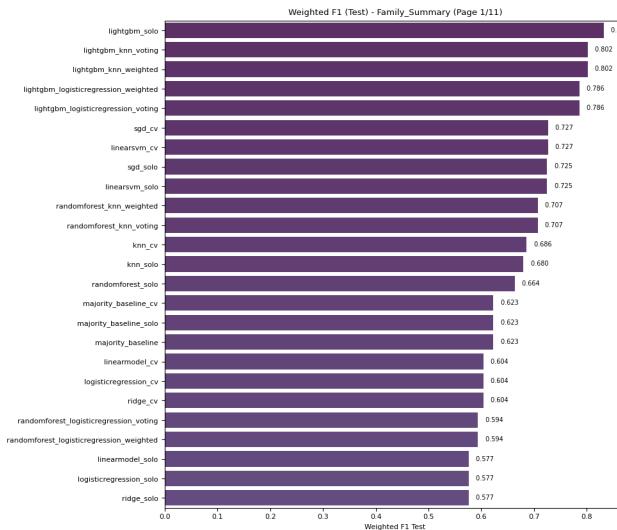
Plot: f1_test_double_only_page_4.png (see [HTML](#) for interactive version)



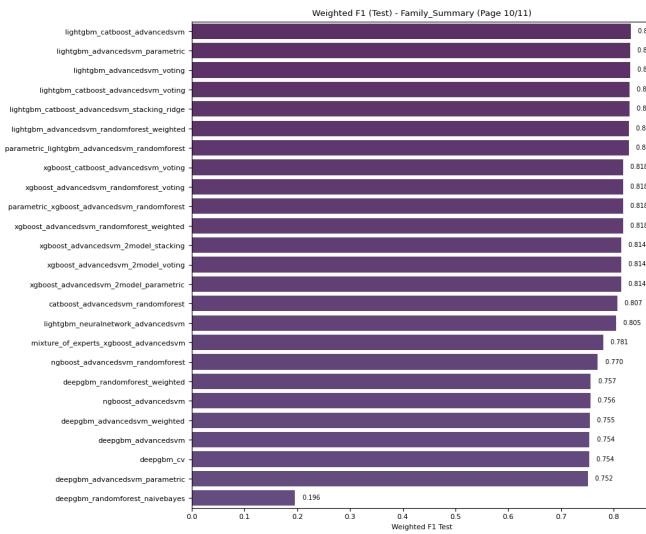
Plot: f1_test_double_only_page_5.png (see [HTML](#) for interactive version)



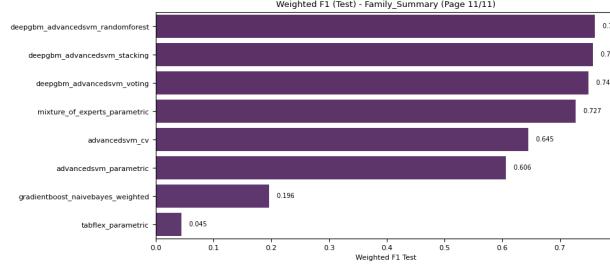
Plot: f1_test_family_summary_page_1.png (see [HTML](#) for interactive version)



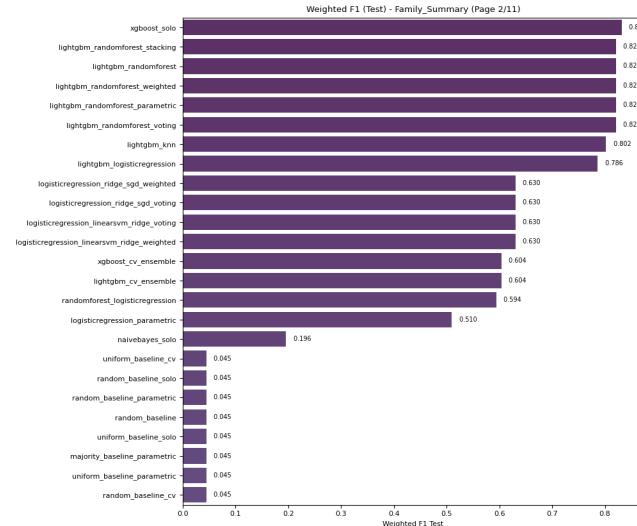
Plot: f1_test_family_summary_page_10.png (see [HTML](#) for interactive version)



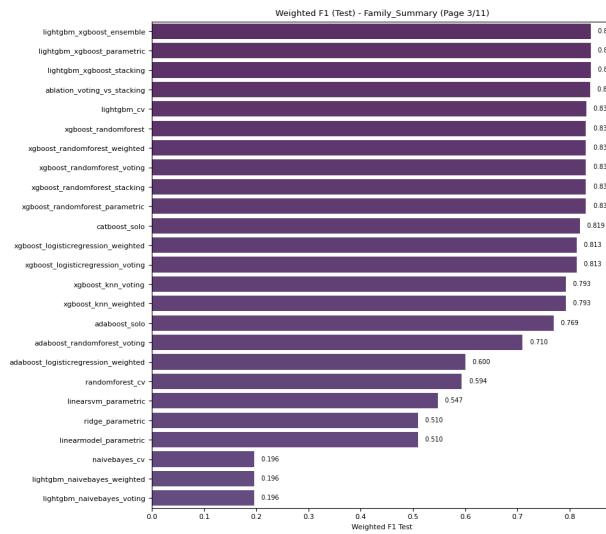
Plot: f1_test_family_summary_page_11.png (see [HTML](#) for interactive version)



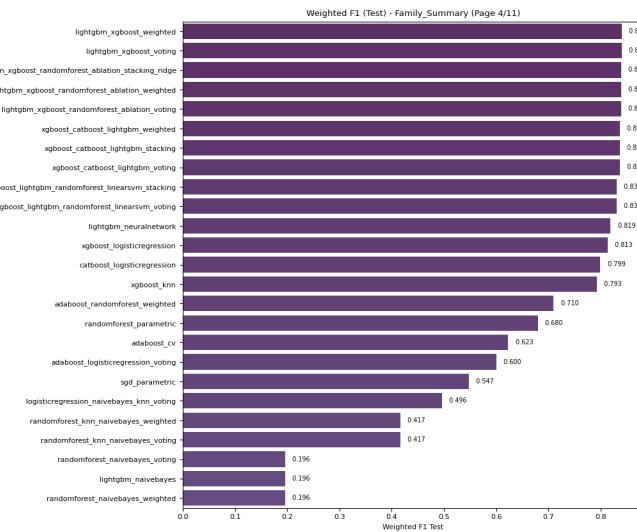
Plot: f1_test_family_summary_page_2.png (see [HTML](#) for interactive version)



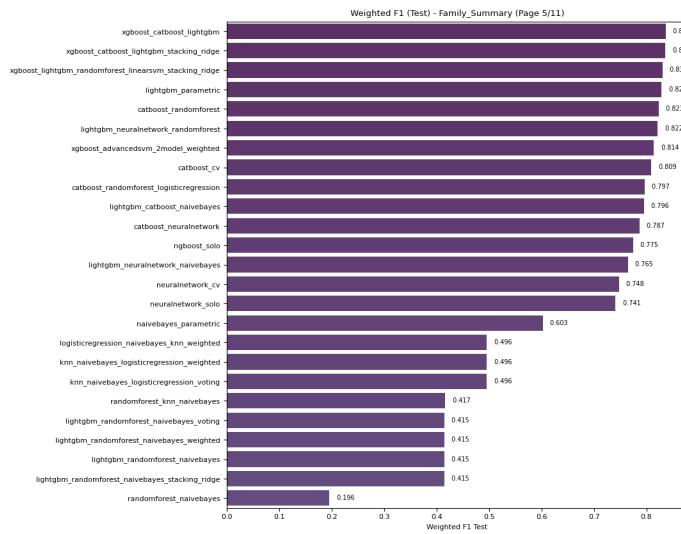
Plot: f1_test_family_summary_page_3.png (see [HTML](#) for interactive version)



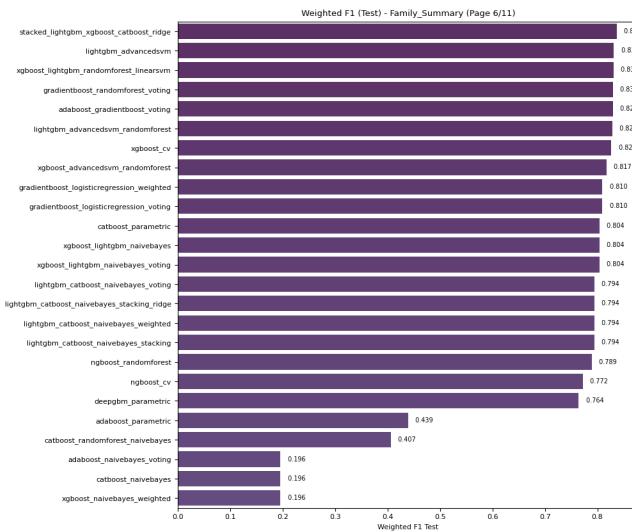
Plot: f1_test_family_summary_page_4.png (see HTML for interactive version)



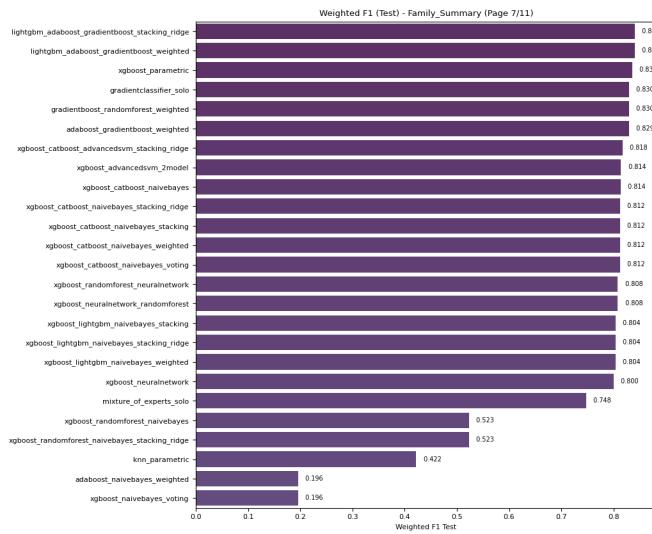
Plot: f1_test_family_summary_page_5.png (see HTML for interactive version)



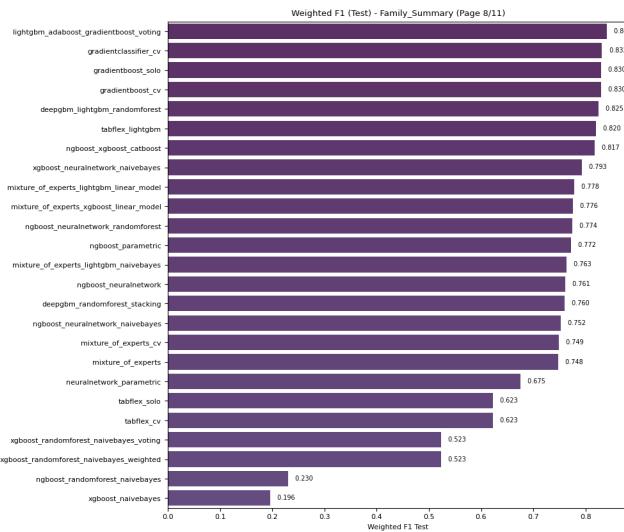
Plot: f1_test_family_summary_page_6.png (see HTML for interactive version)



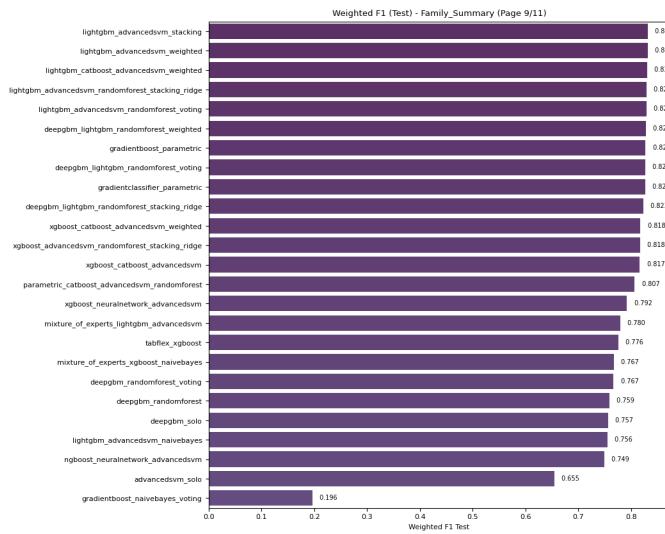
Plot: f1_test_family_summary_page_7.png (see HTML for interactive version)



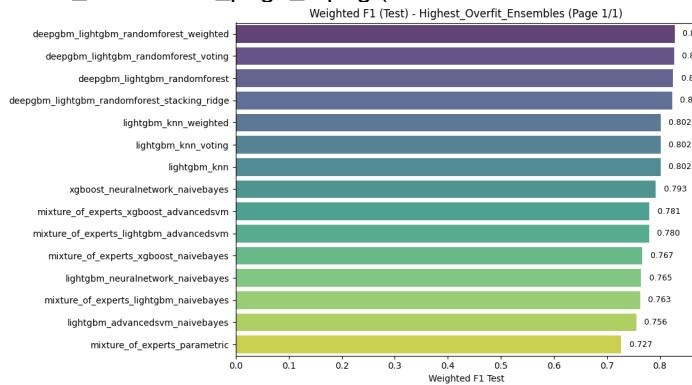
Plot: f1_test_family_summary_page_8.png (see HTML for interactive version)



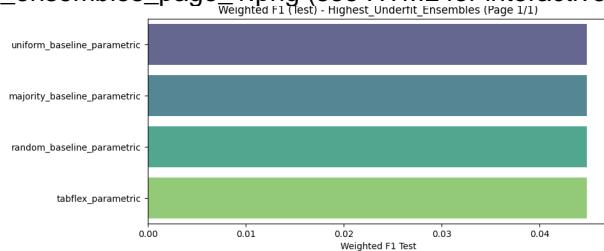
Plot: f1_test_family_summary_page_9.png (see HTML for interactive version)



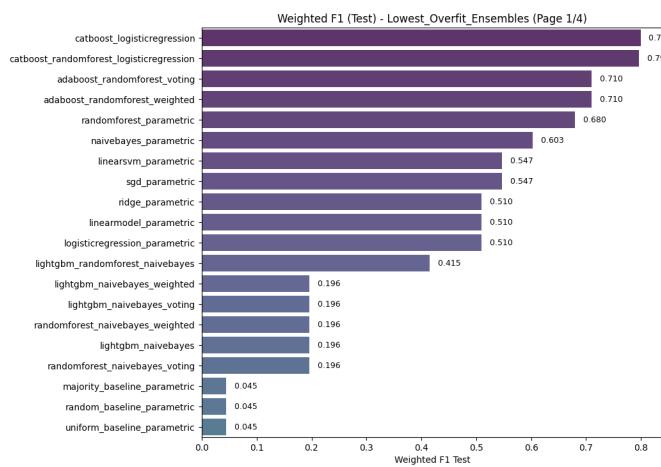
Plot: f1_test_highest_overfit_ensembles_page_1.png (see HTML for interactive version)



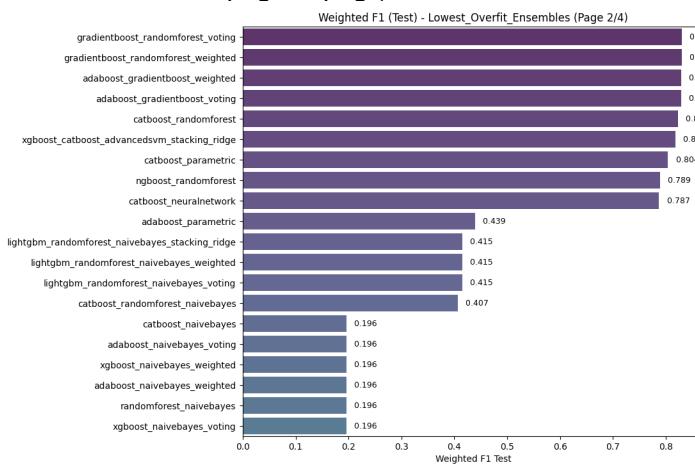
Plot: f1_test_highest_underfit_ensembles_page_1.png (see HTML for interactive version)



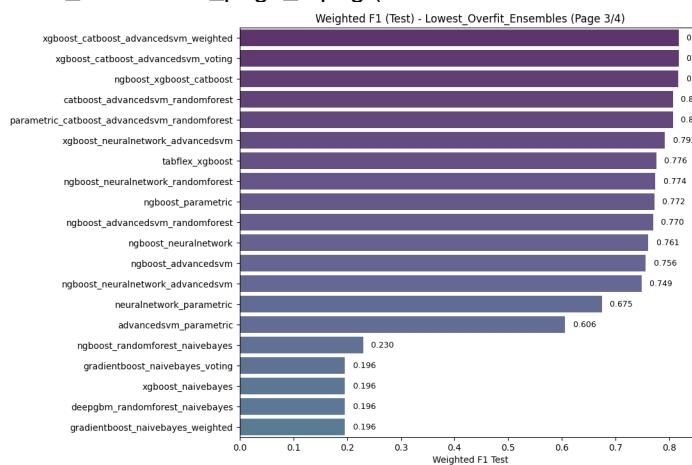
Plot: f1_test_lowest_overfit_ensembles_page_1.png (see HTML for interactive version)



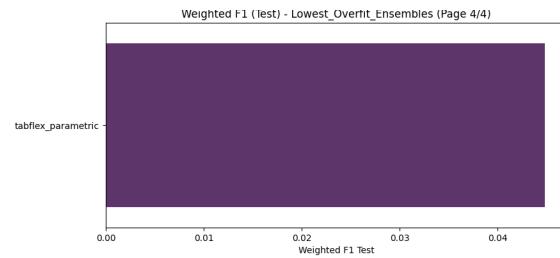
Plot: f1_test_lowest_overfit_ensembles_page_2.png (see HTML for interactive version)



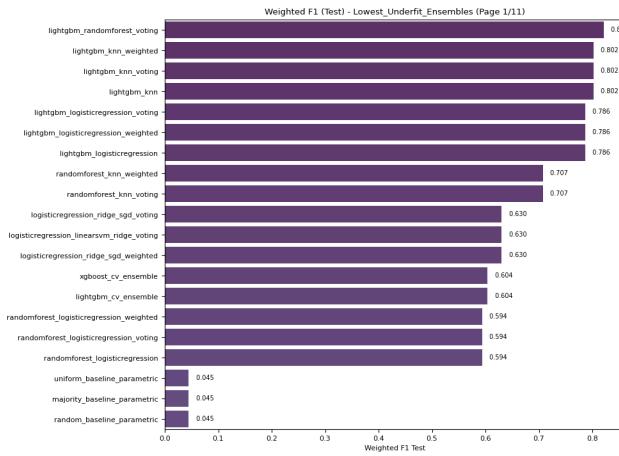
Plot: f1_test_lowest_overfit_ensembles_page_3.png (see HTML for interactive version)



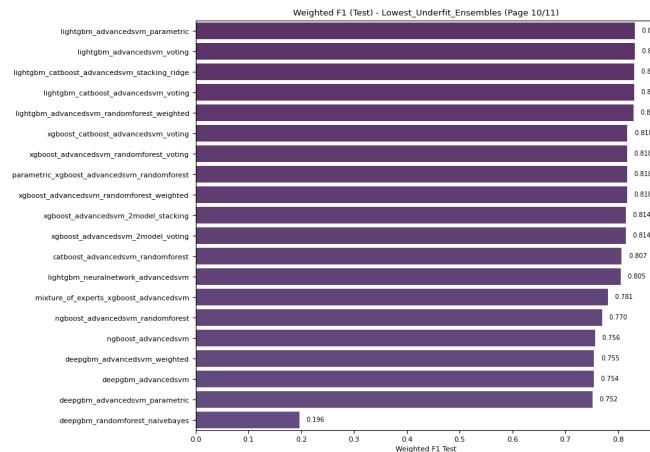
Plot: f1_test_lowest_overfit_ensembles_page_4.png (see HTML for interactive version)



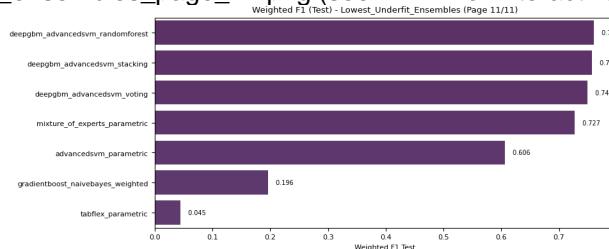
Plot: f1_test_lowest_underfit_ensembles_page_1.png (see [HTML](#) for interactive version)



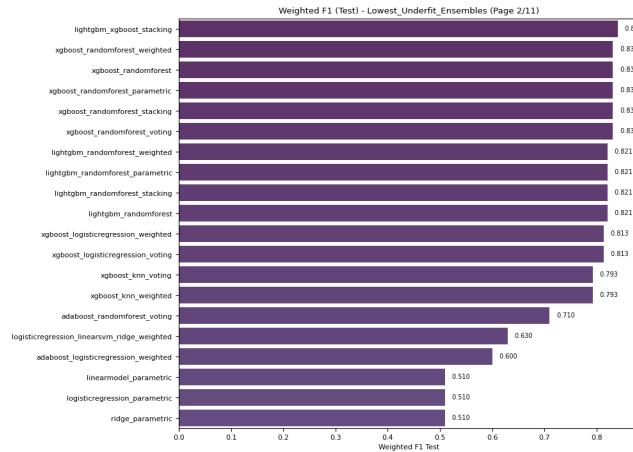
Plot: f1_test_lowest_underfit_ensembles_page_10.png (see [HTML](#) for interactive version)



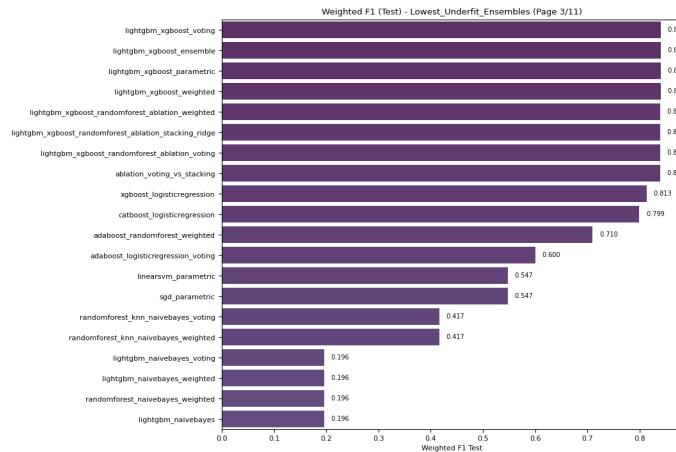
Plot: f1_test_lowest_underfit_ensembles_page_11.png (see [HTML](#) for interactive version)



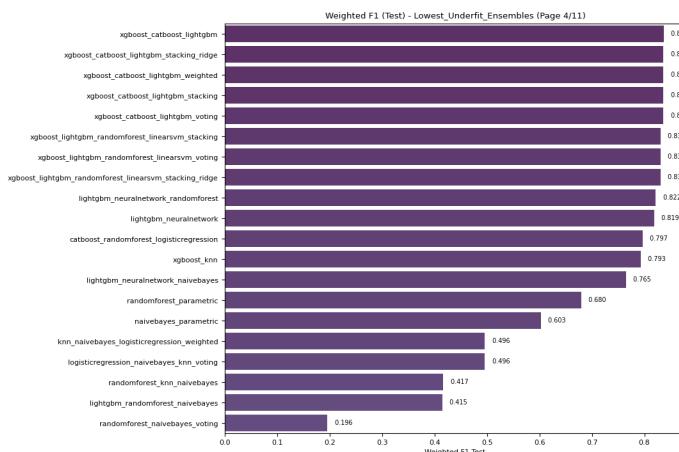
Plot: f1_test_lowest_underfit_ensembles_page_2.png (see [HTML](#) for interactive version)



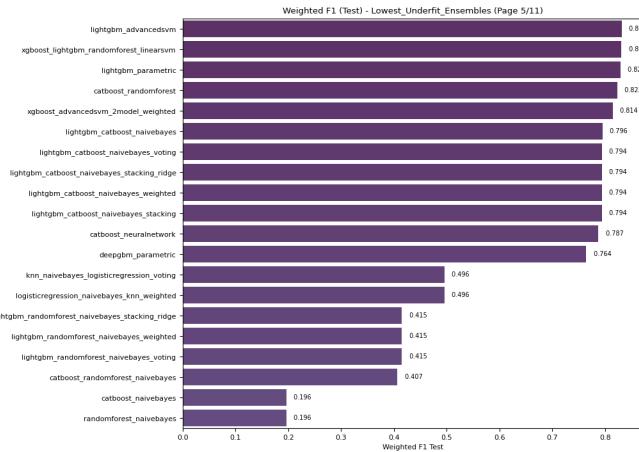
Plot: f1_test_lowest_underfit_ensembles_page_3.png (see [HTML](#) for interactive version)



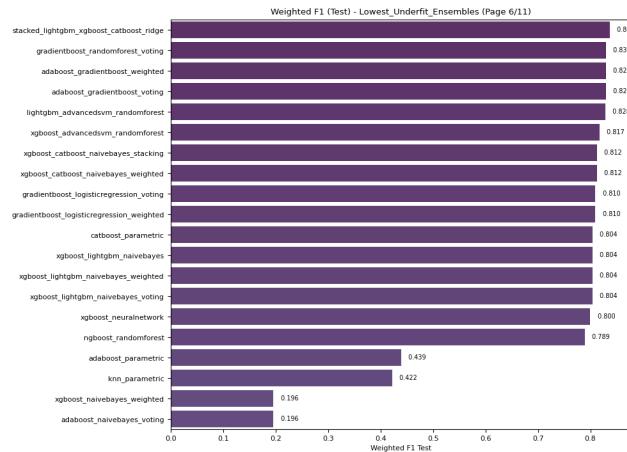
Plot: f1_test_lowest_underfit_ensembles_page_4.png (see [HTML](#) for interactive version)



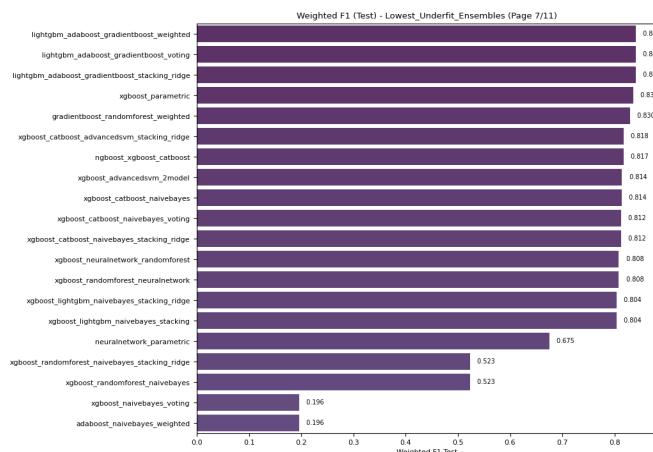
Plot: f1_test_lowest_underfit_ensembles_page_5.png (see [HTML](#) for interactive version)



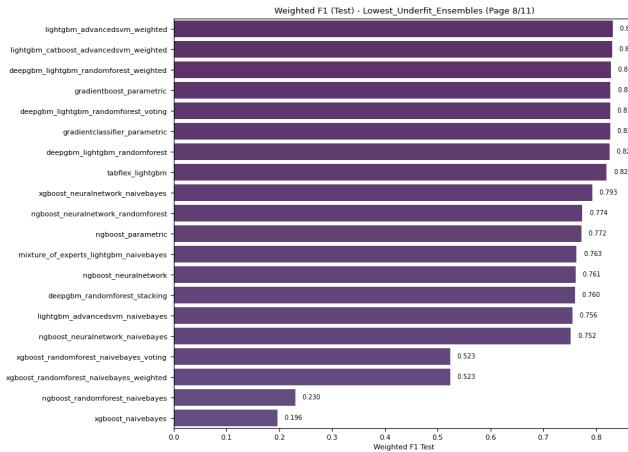
Plot: f1_test_lowest_underfit_ensembles_page_6.png (see [HTML](#) for interactive version)



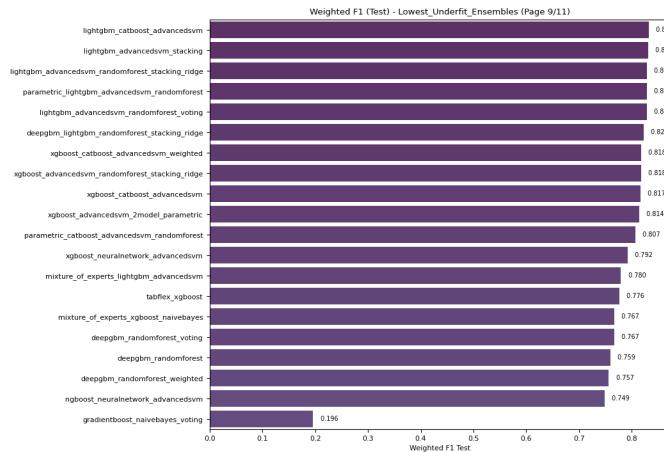
Plot: f1_test_lowest_underfit_ensembles_page_7.png (see [HTML](#) for interactive version)



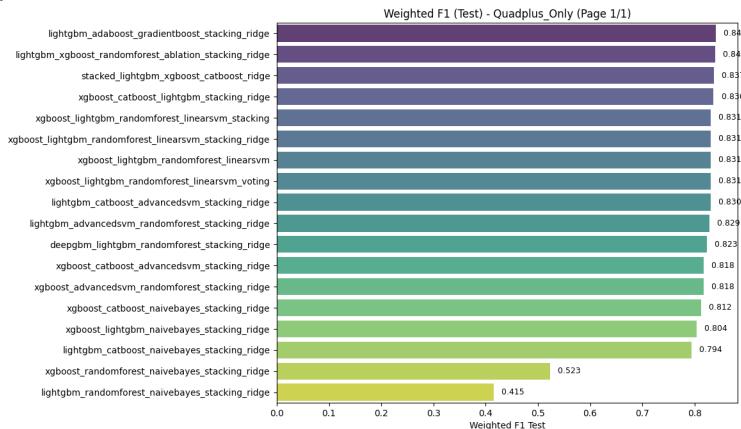
Plot: f1_test_lowest_underfit_ensembles_page_8.png (see [HTML](#) for interactive version)



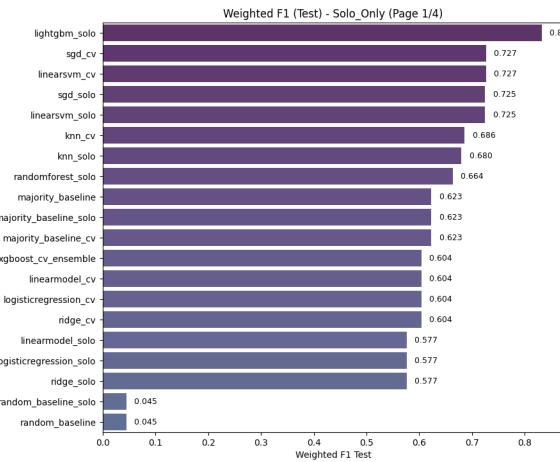
Plot: f1_test_lowest_underfit_ensembles_page_9.png (see HTML for interactive version)



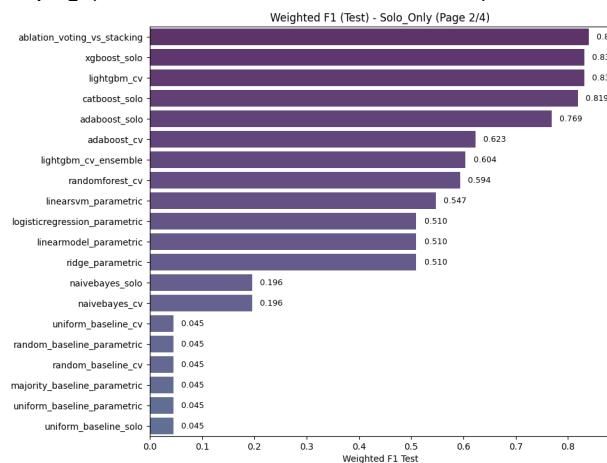
Plot: f1_test_quadplus_only_page_1.png (see HTML for interactive version)



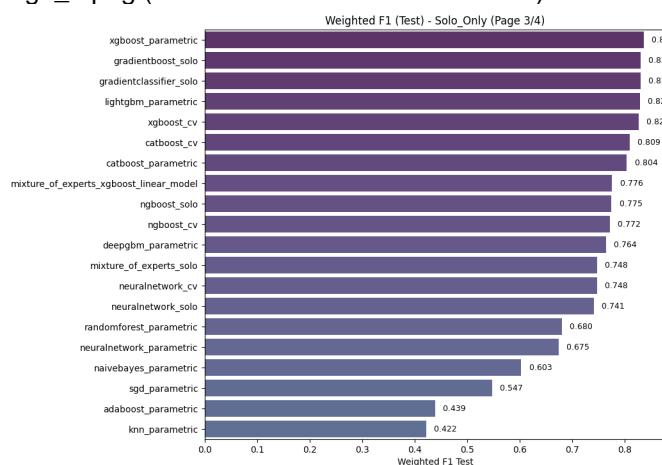
Plot: f1_test_solo_only_page_1.png (see HTML for interactive version)



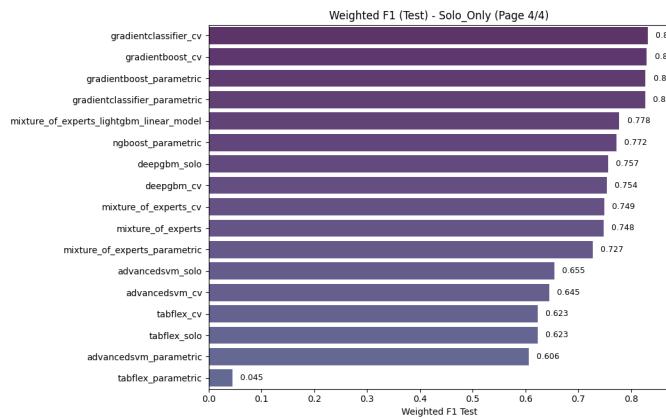
Plot: f1_test_solo_only_page_2.png (see [HTML](#) for interactive version)



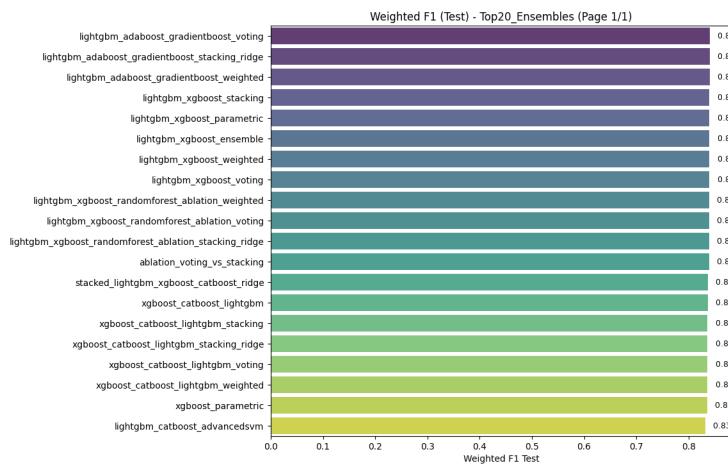
Plot: f1_test_solo_only_page_3.png (see [HTML](#) for interactive version)



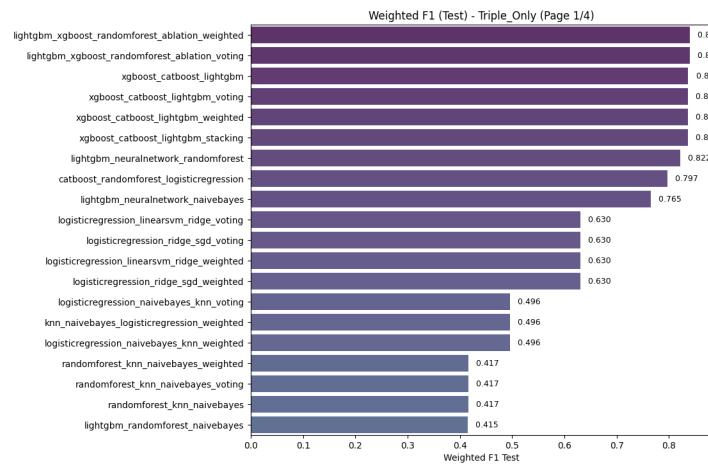
Plot: f1_test_solo_only_page_4.png (see [HTML](#) for interactive version)



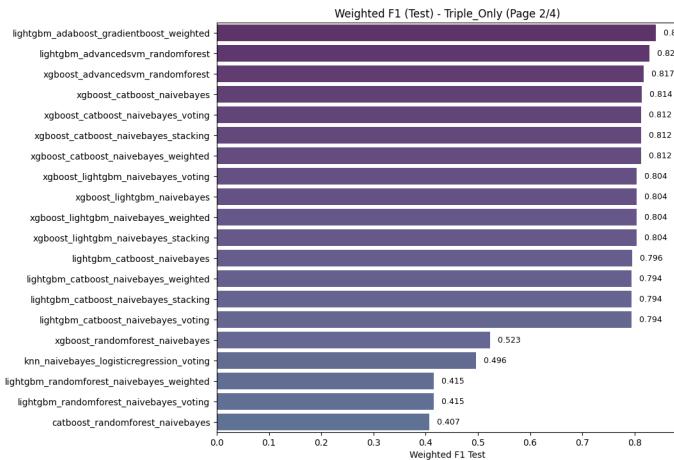
Plot: f1_test_top20_ensembles_page_1.png (see [HTML](#) for interactive version)



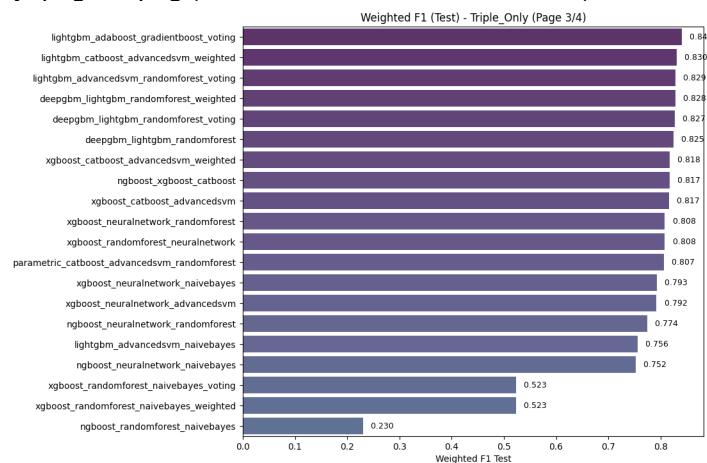
Plot: f1_test_triple_only_page_1.png (see [HTML](#) for interactive version)



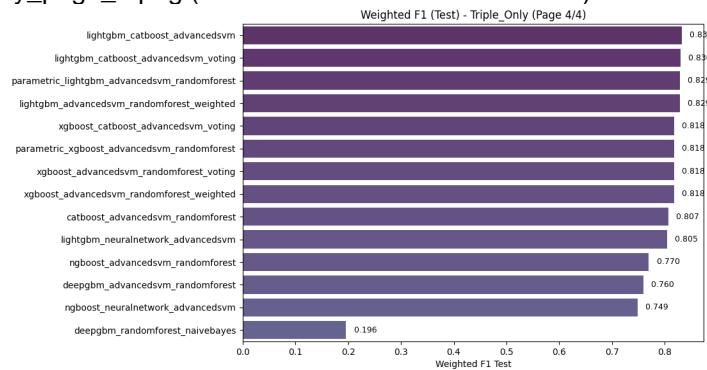
Plot: f1_test_triple_only_page_2.png (see [HTML](#) for interactive version)



Plot: f1_test_triple_only_page_3.png (see [HTML](#) for interactive version)



Plot: f1_test_triple_only_page_4.png (see [HTML](#) for interactive version)



Plot: f1_test_worst20_ensembles_page_1.png (see [HTML](#) for interactive version)

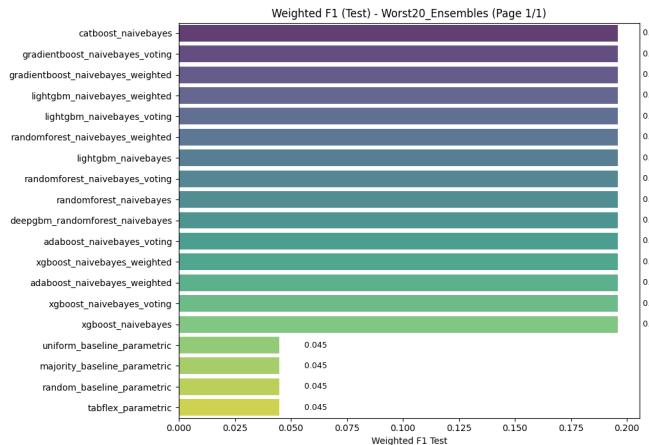


Table Paths

top20_ensembles	analysis_res\top20_ensembles.csv
worst20_ensembles	analysis_res\worst20_ensembles.csv
highest_overfit_ensembles	analysis_res\highest_overfit_ensembles.csv
lowest_overfit_ensembles	analysis_res\lowest_overfit_ensembles.csv
highest_underfit_ensembles	analysis_res\highest_underfit_ensembles.csv
lowest_underfit_ensembles	analysis_res\lowest_underfit_ensembles.csv
all_voting_ensembles	analysis_res\all_voting_ensembles.csv
all_stacking_ensembles	analysis_res\all_stacking_ensembles.csv
all_parametric_ensembles	analysis_res\all_parametric_ensembles.csv
all_cv_ensembles	analysis_res\all_cv_ensembles.csv
all_weighted_ensembles	analysis_res\all_weighted_ensembles.csv
compare_solo	analysis_res\compare_solo_only.csv
compare_double	analysis_res\compare_double_only.csv
compare_triple	analysis_res\compare_triple_only.csv
compare_quadplus	analysis_res\compare_quadplus_only.csv
family_summary	analysis_res\family_summary.csv
complexity	analysis_res\complexity_vs_perf.csv

Appendices

Full datasets in CSVs; validate on external data (e.g., UCI) for generalizability.