

Supplementary Material for: Comparison of Sampling Methods in Machine Learning Models

Mehdi Basikhasteh

This supplementary document provides detailed simulation results that were omitted from the main manuscript for brevity. All results correspond to the experimental setup described in the main paper.

S1. Detailed Simulation Results

S1.1 Performance under Imbalanced Class Ratio (0.1/0.9)

S1.2 Performance under Balanced Class Ratio (0.4/0.6)

S1.3 Performance applied real datasets

S2. Additional Figures

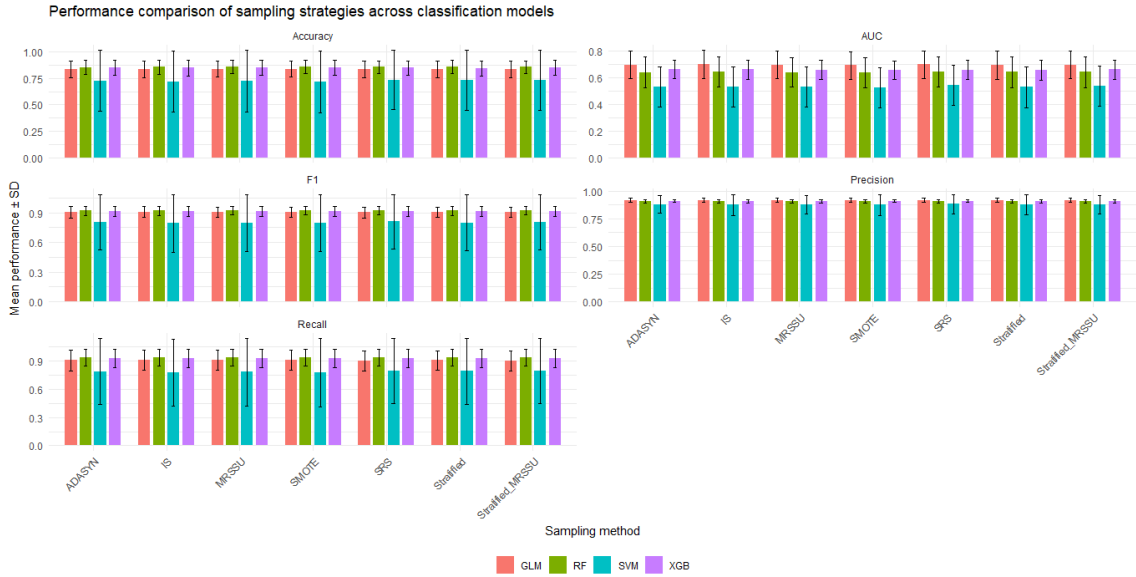


Figure 1: Comparison of sampling methods for class ratio (0.1, 0.9) and sample size 100.

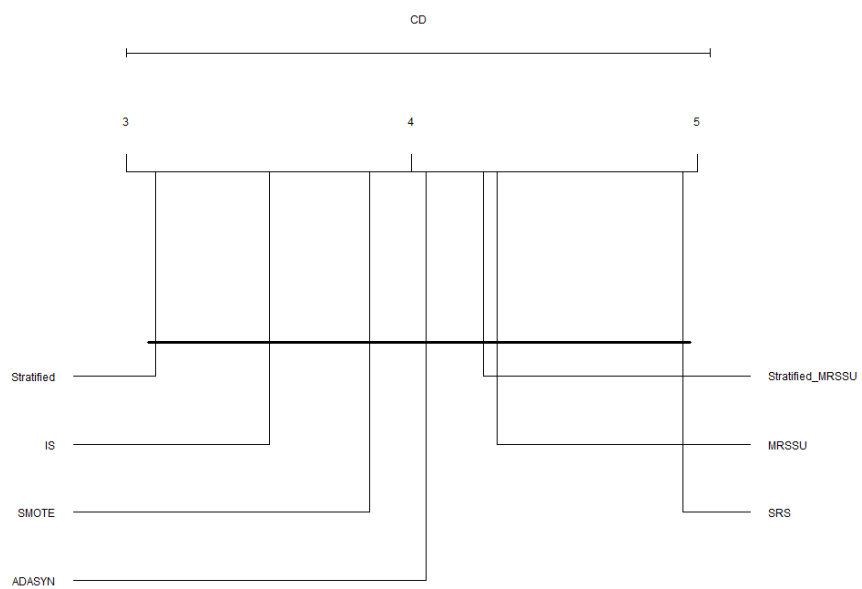


Figure 2: Critical Difference Diagram for Class Ratio (0.1, 0.9) and Sample Size 100

Table 1: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods under scenario $n = 100$, class ratio 0.1/0.9.

Method	Metric	GLM		RF		SVM		XGB	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
ADASYN	Accuracy	0.8361	0.0808	0.8556	0.0639	0.7290	0.2849	0.8503	0.0690
	Precision	0.9170	0.0187	0.9105	0.0148	0.8812	0.0808	0.9119	0.0147
	Recall	0.9015	0.1097	0.9326	0.0890	0.7848	0.3496	0.9247	0.0969
	AUC	0.6937	0.1053	0.6374	0.1134	0.5294	0.1504	0.6590	0.0699
	F1-Score	0.9046	0.0567	0.9185	0.0437	0.8036	0.2823	0.9147	0.0480
IS	Accuracy	0.8381	0.0782	0.8560	0.0638	0.7198	0.2908	0.8481	0.0719
	Precision	0.9170	0.0188	0.9105	0.0151	0.8748	0.0969	0.9113	0.0149
	Recall	0.9040	0.1074	0.9331	0.0892	0.7743	0.3563	0.9227	0.0999
	AUC	0.6959	0.1062	0.6413	0.1128	0.5295	0.1507	0.6593	0.0715
	F1-Score	0.9060	0.0547	0.9187	0.0434	0.7929	0.2891	0.9133	0.0503
MRSSU	Accuracy	0.8380	0.0772	0.8577	0.0634	0.7235	0.2906	0.8498	0.0712
	Precision	0.9172	0.0182	0.9102	0.0143	0.8774	0.0853	0.9111	0.0143
	Recall	0.9035	0.1054	0.9356	0.0886	0.7787	0.3557	0.9250	0.0985
	AUC	0.6942	0.1011	0.6367	0.1090	0.5292	0.1492	0.6549	0.0715
	F1-Score	0.9061	0.0537	0.9198	0.0435	0.7950	0.2908	0.9144	0.0495
SMOTE	Accuracy	0.8390	0.0751	0.8570	0.0628	0.7179	0.2956	0.8498	0.0697
	Precision	0.9166	0.0182	0.9107	0.0146	0.8768	0.0943	0.9115	0.0144
	Recall	0.9055	0.1028	0.9342	0.0874	0.7714	0.3622	0.9246	0.0968
	AUC	0.6890	0.1044	0.6356	0.1126	0.5231	0.1474	0.6539	0.0713
	F1-Score	0.9070	0.0520	0.9194	0.0427	0.7973	0.2889	0.9145	0.0482
SRS	Accuracy	0.8357	0.0784	0.8571	0.0613	0.7348	0.2811	0.8490	0.0696
	Precision	0.9171	0.0182	0.9108	0.0146	0.8821	0.0840	0.9114	0.0142
	Recall	0.9007	0.1062	0.9342	0.0860	0.7911	0.3446	0.9236	0.0964
	AUC	0.6952	0.1043	0.6412	0.1099	0.5406	0.1490	0.6558	0.0719
	F1-Score	0.9046	0.0544	0.9196	0.0418	0.8110	0.2737	0.9140	0.0484
Stratified	Accuracy	0.8361	0.0773	0.8564	0.0635	0.7299	0.2842	0.8476	0.0711
	Precision	0.9167	0.0186	0.9105	0.0152	0.8758	0.0898	0.9113	0.0147
	Recall	0.9017	0.1052	0.9336	0.0886	0.7871	0.3473	0.9219	0.0983
	AUC	0.6922	0.1071	0.6388	0.1150	0.5269	0.1533	0.6563	0.0751
	F1-Score	0.9050	0.0535	0.9190	0.0435	0.7982	0.2857	0.9130	0.0491
SM	Accuracy	0.8358	0.0776	0.8562	0.0622	0.7332	0.2819	0.8485	0.0710
	Precision	0.9172	0.0185	0.9107	0.0151	0.8783	0.0861	0.9117	0.0149
	Recall	0.9009	0.1061	0.9332	0.0874	0.7908	0.3446	0.9227	0.0990
	AUC	0.6937	0.1046	0.6400	0.1148	0.5334	0.1500	0.6584	0.0710
	F1-Score	0.9047	0.0538	0.9190	0.0424	0.8026	0.2806	0.9135	0.0494

Table 2: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods under scenario $n = 180$, class ratio 0.1/0.9.

Method	Metric	GLM		RF		SVM		XGB	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
ADASYN	Accuracy	0.8398	0.0783	0.8450	0.0761	0.7098	0.2921	0.8524	0.0682
	Precision	0.9168	0.0186	0.9097	0.0165	0.8676	0.1109	0.9117	0.0153
	Recall	0.9063	0.1075	0.9198	0.0970	0.7617	0.3565	0.9273	0.0954
	AUC	0.7109	0.0987	0.6455	0.1269	0.5285	0.1546	0.6768	0.0686
	F1-Score	0.9072	0.0540	0.9117	0.0507	0.7811	0.2973	0.9161	0.0467
IS	Accuracy	0.8413	0.0769	0.8475	0.0741	0.6988	0.2999	0.8530	0.0685
	Precision	0.9180	0.0181	0.9105	0.0160	0.8687	0.1015	0.9120	0.0147
	Recall	0.9067	0.1059	0.9221	0.0949	0.7469	0.3667	0.9277	0.0948
	AUC	0.7141	0.0954	0.6485	0.1266	0.5311	0.1547	0.6761	0.0713
	F1-Score	0.9081	0.0530	0.9133	0.0490	0.7695	0.3093	0.9165	0.0471
MRSSU	Accuracy	0.8401	0.0785	0.8466	0.0735	0.7251	0.2818	0.8526	0.0662
	Precision	0.9170	0.0183	0.9101	0.0159	0.8710	0.1006	0.9117	0.0146
	Recall	0.9064	0.1077	0.9214	0.0944	0.7805	0.3436	0.9276	0.0931
	AUC	0.7102	0.0992	0.6469	0.1257	0.5346	0.1544	0.6766	0.0659
	F1-Score	0.9073	0.0542	0.9128	0.0486	0.7895	0.2919	0.9164	0.0450
SMOTE	Accuracy	0.8428	0.0770	0.8466	0.0742	0.7180	0.2871	0.8529	0.0673
	Precision	0.9176	0.0186	0.9101	0.0162	0.8738	0.0917	0.9122	0.0148
	Recall	0.9091	0.1061	0.9214	0.0950	0.7713	0.3502	0.9275	0.0944
	AUC	0.7139	0.0970	0.6459	0.1270	0.5370	0.1554	0.6764	0.0680
	F1-Score	0.9090	0.0533	0.9128	0.0490	0.7829	0.2975	0.9165	0.0461
SRS	Accuracy	0.8401	0.0767	0.8478	0.0721	0.7159	0.2885	0.8507	0.0668
	Precision	0.9171	0.0182	0.9102	0.0162	0.8671	0.1115	0.9117	0.0151
	Recall	0.9062	0.1052	0.9228	0.0932	0.7682	0.3523	0.9252	0.0932
	AUC	0.7130	0.0958	0.6488	0.1267	0.5352	0.1541	0.6774	0.0683
	F1-Score	0.9075	0.0526	0.9136	0.0476	0.7839	0.2974	0.9152	0.0452
Stratified	Accuracy	0.8429	0.0770	0.8484	0.0711	0.7295	0.2788	0.8545	0.0650
	Precision	0.9174	0.0184	0.9105	0.0159	0.8738	0.0918	0.9121	0.0146
	Recall	0.9093	0.1058	0.9233	0.0920	0.7859	0.3396	0.9295	0.0915
	AUC	0.7110	0.0986	0.6478	0.1268	0.5358	0.1547	0.6775	0.0684
	F1-Score	0.9091	0.0529	0.9140	0.0467	0.7915	0.2903	0.9176	0.0442
SM	Accuracy	0.8408	0.0772	0.8463	0.0733	0.7163	0.2872	0.8529	0.0661
	Precision	0.9171	0.0189	0.9102	0.0164	0.8710	0.1042	0.9117	0.0154
	Recall	0.9072	0.1064	0.9210	0.0946	0.7691	0.3506	0.9280	0.0929
	AUC	0.7121	0.0985	0.6473	0.1276	0.5367	0.1549	0.6756	0.0726
	F1-Score	0.9078	0.0533	0.9126	0.0485	0.7813	0.2983	0.9165	0.0451

Table 3: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods under scenario $n = 100$, class ratio 0.4/0.6.

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.6717	0.0505	0.6152	0.0917	0.4941	0.1581	0.6302	0.0445
	Precision	0.7278	0.0486	0.6837	0.0806	0.5611	0.1603	0.6944	0.0423
	Recall	0.7311	0.1049	0.6678	0.1265	0.5138	0.2588	0.6939	0.1096
	AUC	0.7250	0.0703	0.6504	0.1187	0.4906	0.2081	0.6662	0.0564
	F1-Score	0.7240	0.0615	0.6711	0.0923	0.5245	0.2058	0.6880	0.0610
IS	Accuracy	0.6746	0.0505	0.6169	0.0907	0.5036	0.1595	0.6302	0.0440
	Precision	0.7299	0.0507	0.6864	0.0794	0.5731	0.1628	0.6944	0.0428
	Recall	0.7373	0.1000	0.6710	0.1248	0.5244	0.2626	0.6985	0.1097
	AUC	0.7272	0.0706	0.6512	0.1176	0.5032	0.2097	0.6661	0.0560
	F1-Score	0.7288	0.0555	0.6741	0.0905	0.5367	0.2056	0.6903	0.0590
MRSSU	Accuracy	0.6721	0.0538	0.6151	0.0921	0.4982	0.1576	0.6278	0.0481
	Precision	0.7279	0.0517	0.6833	0.0804	0.5660	0.1547	0.6923	0.0438
	Recall	0.7322	0.1076	0.6698	0.1306	0.5272	0.2600	0.6912	0.1148
	AUC	0.7246	0.0751	0.6500	0.1176	0.4963	0.2067	0.6633	0.0604
	F1-Score	0.7246	0.0620	0.6714	0.0937	0.5361	0.2007	0.6854	0.0653
SMOTE	Accuracy	0.6707	0.0564	0.6156	0.0929	0.5089	0.1587	0.6269	0.0491
	Precision	0.7286	0.0544	0.6858	0.0821	0.5784	0.1558	0.6935	0.0457
	Recall	0.7283	0.1089	0.6664	0.1269	0.5430	0.2508	0.6876	0.1103
	AUC	0.7228	0.0788	0.6502	0.1200	0.5094	0.2096	0.6616	0.0632
	F1-Score	0.7229	0.0652	0.6714	0.0926	0.5511	0.1952	0.6847	0.0633
SRS	Accuracy	0.6728	0.0528	0.6179	0.0925	0.4887	0.1595	0.6320	0.0446
	Precision	0.7278	0.0504	0.6842	0.0803	0.5574	0.1591	0.6960	0.0407
	Recall	0.7337	0.1054	0.6748	0.1265	0.5103	0.2599	0.6958	0.1140
	AUC	0.7258	0.0725	0.6520	0.1185	0.4892	0.2085	0.6675	0.0558
	F1-Score	0.7257	0.0611	0.6751	0.0927	0.5210	0.2050	0.6891	0.0671
Stratified	Accuracy	0.6728	0.0548	0.6170	0.0918	0.4927	0.1578	0.6296	0.0481
	Precision	0.7289	0.0518	0.6850	0.0799	0.5652	0.1580	0.6938	0.0431
	Recall	0.7327	0.1065	0.6717	0.1256	0.5116	0.2583	0.6937	0.1142
	AUC	0.7248	0.0765	0.6505	0.1191	0.4920	0.2080	0.6651	0.0594
	F1-Score	0.7255	0.0612	0.6738	0.0913	0.5250	0.2030	0.6874	0.0662
SM	Accuracy	0.6718	0.0569	0.6178	0.0918	0.5018	0.1604	0.6296	0.0476
	Precision	0.7270	0.0534	0.6845	0.0794	0.5683	0.1587	0.6934	0.0428
	Recall	0.7339	0.1077	0.6751	0.1259	0.5258	0.2637	0.6939	0.1133
	AUC	0.7241	0.0806	0.6513	0.1183	0.5045	0.2094	0.6646	0.0602
	F1-Score	0.7251	0.0635	0.6754	0.0916	0.5357	0.2067	0.6876	0.0654

Table 4: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods under scenario $n = 180$, class ratio 0.4/0.6.

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.6806	0.0474	0.6210	0.1020	0.4895	0.1703	0.6422	0.0452
	Precision	0.7348	0.0479	0.6870	0.0887	0.5570	0.1641	0.7023	0.0427
	Recall	0.7397	0.0943	0.6713	0.1286	0.5080	0.2444	0.7084	0.1032
	AUC	0.7356	0.0664	0.6564	0.1302	0.4850	0.2175	0.6809	0.0597
	F1-Score	0.7327	0.0514	0.6756	0.1015	0.5219	0.2044	0.7000	0.0567
IS	Accuracy	0.6828	0.0487	0.6237	0.1030	0.5008	0.1738	0.6444	0.0464
	Precision	0.7345	0.0485	0.6887	0.0892	0.5660	0.1688	0.7028	0.0435
	Recall	0.7478	0.0948	0.6778	0.1308	0.5169	0.2535	0.7153	0.1022
	AUC	0.7364	0.0695	0.6587	0.1309	0.5028	0.2207	0.6825	0.0607
	F1-Score	0.7367	0.0522	0.6798	0.1020	0.5304	0.2119	0.7040	0.0576
MRSSU	Accuracy	0.6824	0.0482	0.6218	0.1030	0.5023	0.1713	0.6417	0.0474
	Precision	0.7343	0.0479	0.6877	0.0899	0.5707	0.1672	0.7008	0.0439
	Recall	0.7451	0.1020	0.6717	0.1325	0.5193	0.2461	0.7108	0.1076
	AUC	0.7377	0.0655	0.6574	0.1307	0.5048	0.2189	0.6805	0.0608
	F1-Score	0.7347	0.0567	0.6760	0.1033	0.5333	0.2052	0.7001	0.0620
SMOTE	Accuracy	0.6833	0.0462	0.6235	0.1030	0.4938	0.1729	0.6442	0.0443
	Precision	0.7351	0.0459	0.6879	0.0886	0.5602	0.1676	0.7020	0.0416
	Recall	0.7481	0.0974	0.6788	0.1311	0.5099	0.2515	0.7175	0.0977
	AUC	0.7390	0.0643	0.6581	0.1311	0.4924	0.2198	0.6821	0.0586
	F1-Score	0.7368	0.0522	0.6800	0.1022	0.5226	0.2112	0.7049	0.0530
SRS	Accuracy	0.6809	0.0494	0.6218	0.1023	0.4953	0.1717	0.6420	0.0479
	Precision	0.7339	0.0479	0.6877	0.0887	0.5641	0.1681	0.7016	0.0441
	Recall	0.7427	0.0971	0.6731	0.1299	0.5080	0.2466	0.7098	0.1015
	AUC	0.7356	0.0697	0.6575	0.1314	0.4962	0.2203	0.6806	0.0618
	F1-Score	0.7337	0.0548	0.6769	0.1017	0.5253	0.2067	0.7007	0.0591
Stratified	Accuracy	0.6808	0.0528	0.6223	0.1044	0.4914	0.1735	0.6423	0.0478
	Precision	0.7337	0.0500	0.6876	0.0906	0.5581	0.1693	0.7015	0.0432
	Recall	0.7423	0.1032	0.6740	0.1330	0.5039	0.2484	0.7108	0.1078
	AUC	0.7358	0.0720	0.6582	0.1324	0.4891	0.2209	0.6808	0.0625
	F1-Score	0.7331	0.0604	0.6772	0.1042	0.5207	0.2090	0.7005	0.0632
SM	Accuracy	0.6817	0.0473	0.6221	0.1030	0.5012	0.1723	0.6428	0.0480
	Precision	0.7338	0.0478	0.6864	0.0892	0.5665	0.1653	0.7012	0.0436
	Recall	0.7454	0.0962	0.6754	0.1310	0.5234	0.2467	0.7112	0.1006
	AUC	0.7357	0.0683	0.6577	0.1311	0.5010	0.2204	0.6817	0.0609

Table 5: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Pima Indians Diabetes dataset under the 70/30 split, $n = 100$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.7168	0.0655	0.7017	0.0774	0.5271	0.2096	0.7000	0.0591
	Precision	0.7896	0.0510	0.7676	0.0676	0.6024	0.2042	0.7805	0.0566
	Recall	0.7751	0.1058	0.7871	0.0953	0.5545	0.3059	0.7597	0.0943
	AUC	0.7695	0.0904	0.7408	0.1286	0.5259	0.2726	0.7545	0.0670
	F1-Score	0.7777	0.0689	0.7734	0.0631	0.5625	0.2651	0.7654	0.0562
IS	Accuracy	0.7219	0.0615	0.7074	0.0722	0.5294	0.2105	0.7038	0.0473
	Precision	0.7942	0.0528	0.7773	0.0723	0.6103	0.1852	0.7863	0.0522
	Recall	0.7797	0.0939	0.7862	0.0881	0.5817	0.2932	0.7596	0.0972
	AUC	0.7773	0.0870	0.7468	0.1215	0.5177	0.2717	0.7530	0.0592
	F1-Score	0.7832	0.0555	0.7772	0.0551	0.5828	0.2463	0.7674	0.0475
MRSSU	Accuracy	0.7257	0.0531	0.7066	0.0704	0.5569	0.2118	0.7009	0.0609
	Precision	0.7876	0.0443	0.7669	0.0634	0.6266	0.2030	0.7764	0.0505
	Recall	0.7973	0.0790	0.8003	0.0829	0.5948	0.3024	0.7669	0.1014
	AUC	0.7726	0.0854	0.7395	0.1217	0.5544	0.2616	0.7481	0.0717
	F1-Score	0.7899	0.0473	0.7798	0.0531	0.5974	0.2632	0.7672	0.0576
SMOTE	Accuracy	0.7160	0.0590	0.6991	0.0782	0.5300	0.2072	0.6934	0.0575
	Precision	0.7799	0.0506	0.7635	0.0682	0.6103	0.1955	0.7743	0.0581
	Recall	0.7911	0.0825	0.7883	0.0839	0.5508	0.3030	0.7584	0.0929
	AUC	0.7604	0.0900	0.7360	0.1313	0.5252	0.2630	0.7441	0.0715
	F1-Score	0.7828	0.0510	0.7729	0.0605	0.5626	0.2648	0.7613	0.0536
SRS	Accuracy	0.7196	0.0691	0.7075	0.0794	0.5323	0.2150	0.7037	0.0528
	Precision	0.7815	0.0508	0.7654	0.0672	0.6062	0.2060	0.7749	0.0474
	Recall	0.7928	0.0962	0.8039	0.0903	0.5539	0.3141	0.7775	0.0975
	AUC	0.7617	0.0993	0.7419	0.1354	0.5298	0.2688	0.7492	0.0697
	F1-Score	0.7844	0.0628	0.7809	0.0612	0.5623	0.2732	0.7714	0.0510
Stratified	Accuracy	0.7162	0.0608	0.7002	0.0762	0.5263	0.2111	0.6985	0.0521
	Precision	0.7833	0.0435	0.7630	0.0687	0.5992	0.2004	0.7797	0.0519
	Recall	0.7837	0.1023	0.7933	0.0863	0.5592	0.3089	0.7598	0.1064
	AUC	0.7644	0.0832	0.7381	0.1293	0.5188	0.2678	0.7485	0.0670
	F1-Score	0.7797	0.0613	0.7746	0.0580	0.5695	0.2603	0.7635	0.0557
SM	Accuracy	0.7256	0.0645	0.7188	0.0707	0.5368	0.2153	0.7108	0.0556
	Precision	0.7934	0.0511	0.7795	0.0685	0.6139	0.2013	0.7893	0.0531
	Recall	0.7867	0.0954	0.8047	0.0781	0.5772	0.2970	0.7661	0.0926
	AUC	0.7738	0.0960	0.7533	0.1336	0.5335	0.2747	0.7620	0.0719
	F1-Score	0.7867	0.0600	0.7884	0.0513	0.5831	0.2568	0.7734	0.0534

Table 6: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Pima Indians Diabetes dataset under the 70/30 split, $n = 180$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.7282	0.0635	0.7043	0.1007	0.5118	0.2255	0.7191	0.0473
	Precision	0.7851	0.0456	0.7670	0.0798	0.5889	0.2081	0.7878	0.0453
	Recall	0.8049	0.1002	0.7883	0.1040	0.5355	0.3104	0.7867	0.0906
	AUC	0.7803	0.0895	0.7386	0.1524	0.5052	0.2867	0.7709	0.0540
	F1-Score	0.7916	0.0615	0.7751	0.0816	0.5478	0.2699	0.7827	0.0521
IS	Accuracy	0.7301	0.0666	0.7064	0.1030	0.5240	0.2226	0.7129	0.0614
	Precision	0.7940	0.0469	0.7690	0.0821	0.6047	0.2049	0.7938	0.0547
	Recall	0.7939	0.1070	0.7899	0.1079	0.5509	0.2998	0.7657	0.1073
	AUC	0.7877	0.0875	0.7417	0.1598	0.5177	0.2851	0.7674	0.0823
	F1-Score	0.7899	0.0673	0.7765	0.0837	0.5647	0.2620	0.7734	0.0633
MRSSU	Accuracy	0.7288	0.0702	0.7082	0.1019	0.5518	0.2216	0.7153	0.0569
	Precision	0.7925	0.0494	0.7741	0.0833	0.6330	0.2020	0.7908	0.0561
	Recall	0.7934	0.1028	0.7859	0.0948	0.5853	0.2870	0.7743	0.0809
	AUC	0.7849	0.0950	0.7439	0.1672	0.5518	0.2878	0.7665	0.0852
	F1-Score	0.7896	0.0648	0.7778	0.0784	0.5982	0.2534	0.7789	0.0475
SMOTE	Accuracy	0.7380	0.0450	0.7060	0.0980	0.5473	0.2191	0.7201	0.0454
	Precision	0.7972	0.0443	0.7672	0.0788	0.6246	0.1923	0.7901	0.0482
	Recall	0.8081	0.0734	0.7927	0.0978	0.5935	0.2930	0.7849	0.0781
	AUC	0.7938	0.0721	0.7431	0.1586	0.5445	0.2813	0.7757	0.0558
	F1-Score	0.7998	0.0389	0.7775	0.0774	0.5979	0.2509	0.7840	0.0416
SRS	Accuracy	0.7276	0.0529	0.6930	0.0913	0.5183	0.2139	0.7065	0.0530
	Precision	0.7848	0.0444	0.7575	0.0708	0.6012	0.1916	0.7822	0.0462
	Recall	0.8063	0.0839	0.7812	0.0972	0.5468	0.2949	0.7673	0.0896
	AUC	0.7784	0.0736	0.7332	0.1449	0.5112	0.2717	0.7569	0.0691
	F1-Score	0.7926	0.0488	0.7670	0.0737	0.5605	0.2541	0.7711	0.0516
Stratified	Accuracy	0.7207	0.0788	0.7068	0.0985	0.5002	0.2227	0.7188	0.0492
	Precision	0.7835	0.0620	0.7661	0.0799	0.5784	0.2031	0.7885	0.0501
	Recall	0.7907	0.0976	0.7969	0.1011	0.5249	0.3104	0.7849	0.0792
	AUC	0.7673	0.1125	0.7443	0.1548	0.4925	0.2843	0.7689	0.0706
	F1-Score	0.7848	0.0707	0.7787	0.0790	0.5366	0.2685	0.7832	0.0455
SM	Accuracy	0.7307	0.0546	0.7040	0.0993	0.5224	0.2247	0.7170	0.0511
	Precision	0.7917	0.0455	0.7654	0.0782	0.5987	0.2163	0.7869	0.0508
	Recall	0.8016	0.0876	0.7914	0.1009	0.5381	0.3090	0.7839	0.0828
	AUC	0.7877	0.0796	0.7415	0.1539	0.5214	0.2829	0.7662	0.0712
	F1-Score	0.7933	0.0504	0.7759	0.0787	0.5537	0.2736	0.7818	0.0458

Table 7: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Pima Indians Diabetes dataset using 5-fold cross-validation, $n = 49$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.7005	0.0426	0.7023	0.0459	0.5314	0.0999	0.6860	0.0359
	Precision	0.7814	0.0371	0.7692	0.0533	0.6091	0.0918	0.7706	0.0372
	Recall	0.7547	0.0625	0.7917	0.0626	0.5826	0.1912	0.7496	0.0607
	AUC	0.7432	0.0708	0.7403	0.0966	0.5206	0.1222	0.7300	0.0482
	F1-Score	0.7663	0.0393	0.7771	0.0318	0.5944	0.1360	0.7577	0.0315
IS	Accuracy	0.6958	0.0494	0.7008	0.0509	0.5396	0.1106	0.6846	0.0369
	Precision	0.7755	0.0408	0.7660	0.0565	0.6236	0.0921	0.7680	0.0408
	Recall	0.7554	0.0694	0.7952	0.0685	0.5948	0.1950	0.7513	0.0678
	AUC	0.7373	0.0736	0.7408	0.1020	0.5318	0.1363	0.7283	0.0521
	F1-Score	0.7637	0.0449	0.7769	0.0368	0.6159	0.1288	0.7569	0.0341
MRSSU	Accuracy	0.6981	0.0495	0.7006	0.0493	0.5168	0.1069	0.6848	0.0372
	Precision	0.7772	0.0367	0.7648	0.0539	0.5919	0.1098	0.7684	0.0388
	Recall	0.7560	0.0724	0.7958	0.0603	0.5531	0.2074	0.7485	0.0675
	AUC	0.7396	0.0726	0.7397	0.0973	0.5047	0.1286	0.7286	0.0486
	F1-Score	0.7649	0.0470	0.7771	0.0346	0.5728	0.1507	0.7559	0.0370
SMOTE	Accuracy	0.6966	0.0524	0.7013	0.0509	0.5326	0.1049	0.6866	0.0363
	Precision	0.7787	0.0398	0.7653	0.0533	0.6131	0.0980	0.7672	0.0372
	Recall	0.7499	0.0723	0.7954	0.0594	0.5756	0.1977	0.7554	0.0625
	AUC	0.7365	0.0791	0.7382	0.0998	0.5203	0.1299	0.7287	0.0490
	F1-Score	0.7625	0.0486	0.7775	0.0358	0.6000	0.1393	0.7592	0.0337
SRS	Accuracy	0.6981	0.0467	0.7002	0.0477	0.5394	0.1001	0.6847	0.0354
	Precision	0.7765	0.0369	0.7654	0.0514	0.6138	0.0973	0.7665	0.0358
	Recall	0.7579	0.0662	0.7924	0.0626	0.5933	0.1855	0.7541	0.0620
	AUC	0.7402	0.0728	0.7386	0.0944	0.5339	0.1233	0.7262	0.0519
	F1-Score	0.7656	0.0430	0.7759	0.0348	0.6011	0.1364	0.7580	0.0315
Stratified	Accuracy	0.6998	0.0410	0.7022	0.0478	0.5428	0.0961	0.6885	0.0334
	Precision	0.7773	0.0371	0.7671	0.0533	0.6227	0.0861	0.7716	0.0370
	Recall	0.7613	0.0619	0.7946	0.0655	0.6044	0.1830	0.7533	0.0619
	AUC	0.7418	0.0684	0.7432	0.0962	0.5351	0.1212	0.7332	0.0499
	F1-Score	0.7676	0.0369	0.7775	0.0348	0.6146	0.1185	0.7600	0.0305
SM	Accuracy	0.6984	0.0500	0.7023	0.0457	0.5304	0.0979	0.6844	0.0396
	Precision	0.7787	0.0378	0.7671	0.0516	0.6022	0.1070	0.7711	0.0374
	Recall	0.7536	0.0686	0.7946	0.0605	0.5626	0.1939	0.7431	0.0676
	AUC	0.7411	0.0751	0.7426	0.0933	0.5242	0.1186	0.7306	0.0497
	F1-Score	0.7646	0.0463	0.7778	0.0325	0.5797	0.1504	0.7546	0.0387

Table 8: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Pima Indians Diabetes dataset using 5-fold cross-validation, $n = 100$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.7228	0.0462	0.7053	0.0730	0.5121	0.0923	0.7042	0.0341
	Precision	0.7871	0.0372	0.7687	0.0645	0.5922	0.0909	0.7771	0.0400
	Recall	0.7911	0.0661	0.7928	0.0687	0.5340	0.1479	0.7750	0.0521
	AUC	0.7739	0.0745	0.7414	0.1315	0.5038	0.1137	0.7532	0.0498
	F1-Score	0.7875	0.0406	0.7786	0.0539	0.5574	0.1192	0.7742	0.0267
IS	Accuracy	0.7187	0.0507	0.7071	0.0749	0.5259	0.0997	0.7039	0.0379
	Precision	0.7860	0.0376	0.7701	0.0656	0.6054	0.1001	0.7786	0.0388
	Recall	0.7841	0.0685	0.7934	0.0655	0.5505	0.1488	0.7708	0.0594
	AUC	0.7712	0.0755	0.7436	0.1346	0.5187	0.1246	0.7542	0.0540
	F1-Score	0.7837	0.0447	0.7799	0.0545	0.5733	0.1241	0.7728	0.0329
MRSSU	Accuracy	0.7221	0.0439	0.7050	0.0716	0.5274	0.0956	0.7046	0.0385
	Precision	0.7851	0.0362	0.7664	0.0632	0.6037	0.0903	0.7787	0.0386
	Recall	0.7955	0.0627	0.7969	0.0636	0.5673	0.1464	0.7722	0.0572
	AUC	0.7714	0.0728	0.7416	0.1326	0.5204	0.1202	0.7536	0.0556
	F1-Score	0.7887	0.0376	0.7795	0.0514	0.5816	0.1189	0.7737	0.0319
SMOTE	Accuracy	0.7235	0.0465	0.7065	0.0727	0.5418	0.1025	0.7025	0.0393
	Precision	0.7877	0.0383	0.7676	0.0643	0.6139	0.0965	0.7757	0.0415
	Recall	0.7914	0.0596	0.7971	0.0699	0.5817	0.1579	0.7753	0.0546
	AUC	0.7693	0.0790	0.7404	0.1335	0.5391	0.1286	0.7495	0.0585
	F1-Score	0.7884	0.0388	0.7800	0.0536	0.5936	0.1255	0.7736	0.0304
SRS	Accuracy	0.7220	0.0493	0.7061	0.0727	0.5319	0.1042	0.7057	0.0350
	Precision	0.7841	0.0375	0.7667	0.0619	0.6054	0.1017	0.7764	0.0362
	Recall	0.7943	0.0660	0.7968	0.0700	0.5661	0.1634	0.7782	0.0560
	AUC	0.7729	0.0731	0.7427	0.1313	0.5257	0.1323	0.7539	0.0525
	F1-Score	0.7878	0.0427	0.7796	0.0540	0.5810	0.1320	0.7755	0.0295
Stratified	Accuracy	0.7206	0.0491	0.7033	0.0753	0.5180	0.1047	0.7032	0.0405
	Precision	0.7874	0.0362	0.7670	0.0661	0.5930	0.1033	0.7779	0.0412
	Recall	0.7865	0.0679	0.7905	0.0693	0.5504	0.1586	0.7699	0.0584
	AUC	0.7693	0.0764	0.7402	0.1315	0.5066	0.1325	0.7495	0.0567
	F1-Score	0.7855	0.0428	0.7767	0.0557	0.5669	0.1304	0.7720	0.0342
SM	Accuracy	0.7209	0.0497	0.7063	0.0650	0.5312	0.0991	0.7020	0.0390
	Precision	0.7855	0.0375	0.7682	0.0597	0.6097	0.1009	0.7766	0.0368
	Recall	0.7900	0.0692	0.7964	0.0618	0.5626	0.1553	0.7694	0.0583
	AUC	0.7708	0.0776	0.7425	0.1238	0.5258	0.1267	0.7505	0.0542
	F1-Score	0.7862	0.0434	0.7800	0.0462	0.5829	0.1276	0.7713	0.0329

Table 9: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Pima Indians Diabetes dataset using 5-fold cross-validation, $n = 180$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.7314	0.0480	0.7055	0.0972	0.5143	0.1001	0.7165	0.0408
	Precision	0.7933	0.0334	0.7699	0.0774	0.5950	0.0950	0.7901	0.0374
	Recall	0.7976	0.0702	0.7860	0.0908	0.5356	0.1329	0.7756	0.0594
	AUC	0.7871	0.0719	0.7428	0.1544	0.5087	0.1322	0.7705	0.0568
	F1-Score	0.7940	0.0428	0.7763	0.0761	0.5614	0.1144	0.7811	0.0346
IS	Accuracy	0.7261	0.0500	0.7026	0.1003	0.5087	0.0961	0.7124	0.0465
	Precision	0.7885	0.0353	0.7678	0.0794	0.5901	0.0876	0.7860	0.0438
	Recall	0.7949	0.0687	0.7823	0.0961	0.5277	0.1326	0.7740	0.0618
	AUC	0.7830	0.0724	0.7409	0.1557	0.5007	0.1253	0.7666	0.0612
	F1-Score	0.7904	0.0437	0.7735	0.0805	0.5545	0.1117	0.7782	0.0390
MRSSU	Accuracy	0.7260	0.0574	0.7011	0.1048	0.5061	0.1026	0.7128	0.0500
	Precision	0.7877	0.0400	0.7649	0.0823	0.5861	0.0926	0.7854	0.0435
	Recall	0.7932	0.0810	0.7840	0.1011	0.5325	0.1355	0.7741	0.0705
	AUC	0.7813	0.0799	0.7407	0.1591	0.4993	0.1311	0.7649	0.0682
	F1-Score	0.7887	0.0527	0.7727	0.0841	0.5557	0.1144	0.7777	0.0454
SMOTE	Accuracy	0.7293	0.0466	0.7010	0.1003	0.5190	0.1122	0.7123	0.0413
	Precision	0.7883	0.0355	0.7662	0.0794	0.5989	0.1088	0.7869	0.0373
	Recall	0.8022	0.0645	0.7815	0.0954	0.5369	0.1454	0.7722	0.0600
	AUC	0.7827	0.0741	0.7375	0.1586	0.5132	0.1474	0.7675	0.0562
	F1-Score	0.7938	0.0395	0.7722	0.0798	0.5639	0.1285	0.7778	0.0349
SRS	Accuracy	0.7314	0.0466	0.7031	0.1000	0.5140	0.0999	0.7142	0.0449
	Precision	0.7911	0.0337	0.7652	0.0793	0.5917	0.0949	0.7843	0.0433
	Recall	0.8015	0.0657	0.7890	0.0916	0.5349	0.1348	0.7810	0.0581
	AUC	0.7855	0.0715	0.7403	0.1576	0.5084	0.1309	0.7664	0.0607
	F1-Score	0.7949	0.0404	0.7756	0.0784	0.5594	0.1153	0.7810	0.0365
Stratified	Accuracy	0.7299	0.0473	0.7040	0.0993	0.5012	0.1200	0.7142	0.0452
	Precision	0.7902	0.0337	0.7682	0.0797	0.5827	0.1142	0.7875	0.0433
	Recall	0.7995	0.0688	0.7846	0.0936	0.5146	0.1581	0.7750	0.0584
	AUC	0.7838	0.0745	0.7432	0.1567	0.4922	0.1547	0.7666	0.0653
	F1-Score	0.7934	0.0419	0.7748	0.0787	0.5437	0.1378	0.7795	0.0366
SM	Accuracy	0.7300	0.0504	0.7033	0.0961	0.5099	0.0965	0.7148	0.0398
	Precision	0.7897	0.0344	0.7663	0.0764	0.5873	0.0909	0.7858	0.0359
	Recall	0.8008	0.0715	0.7866	0.0891	0.5312	0.1312	0.7800	0.0585
	AUC	0.7869	0.0735	0.7414	0.1528	0.5033	0.1249	0.7687	0.0531
	F1-Score	0.7937	0.0447	0.7750	0.0757	0.5556	0.1121	0.7813	0.0339

Table 10: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Breast Cancer dataset under the 70/30 split, $n = 100$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.9367	0.0249	0.9610	0.0142	0.6178	0.4519	0.9372	0.0292
	Precision	0.9451	0.0341	0.9740	0.0215	0.6401	0.4464	0.9571	0.0312
	Recall	0.9602	0.0237	0.9666	0.0157	0.6192	0.4504	0.9475	0.0394
	AUC	0.9593	0.0375	0.9905	0.0051	0.6254	0.4760	0.9806	0.0170
	F1-Score	0.9520	0.0182	0.9700	0.0106	0.6280	0.4498	0.9515	0.0234
IS	Accuracy	0.9305	0.0332	0.9616	0.0172	0.6178	0.4501	0.9429	0.0266
	Precision	0.9391	0.0389	0.9710	0.0230	0.6391	0.4400	0.9577	0.0310
	Recall	0.9574	0.0306	0.9707	0.0150	0.6233	0.4504	0.9562	0.0336
	AUC	0.9442	0.0477	0.9908	0.0044	0.6260	0.4746	0.9833	0.0131
	F1-Score	0.9475	0.0245	0.9706	0.0128	0.6295	0.4469	0.9562	0.0208
MRSSU	Accuracy	0.9251	0.0403	0.9598	0.0155	0.5536	0.4626	0.9409	0.0256
	Precision	0.9371	0.0467	0.9723	0.0226	0.5782	0.4530	0.9584	0.0320
	Recall	0.9514	0.0278	0.9665	0.0117	0.5574	0.4613	0.9521	0.0296
	AUC	0.9369	0.0538	0.9893	0.0037	0.5576	0.4870	0.9807	0.0142
	F1-Score	0.9435	0.0295	0.9692	0.0114	0.5658	0.4589	0.9546	0.0194
SMOTE	Accuracy	0.9271	0.0337	0.9597	0.0178	0.6178	0.4499	0.9375	0.0305
	Precision	0.9333	0.0437	0.9685	0.0260	0.6420	0.4401	0.9510	0.0365
	Recall	0.9592	0.0222	0.9706	0.0139	0.6217	0.4478	0.9554	0.0377
	AUC	0.9426	0.0528	0.9906	0.0036	0.6266	0.4756	0.9821	0.0139
	F1-Score	0.9454	0.0238	0.9693	0.0130	0.6300	0.4456	0.9522	0.0239
SRS	Accuracy	0.9285	0.0372	0.9593	0.0151	0.6075	0.4520	0.9382	0.0319
	Precision	0.9367	0.0408	0.9711	0.0200	0.6293	0.4447	0.9552	0.0311
	Recall	0.9568	0.0317	0.9669	0.0176	0.6108	0.4523	0.9508	0.0348
	AUC	0.9428	0.0495	0.9899	0.0056	0.6164	0.4771	0.9802	0.0160
	F1-Score	0.9460	0.0278	0.9688	0.0115	0.6183	0.4501	0.9525	0.0247
Stratified	Accuracy	0.9312	0.0326	0.9637	0.0178	0.5997	0.4557	0.9441	0.0281
	Precision	0.9400	0.0364	0.9745	0.0211	0.6245	0.4454	0.9596	0.0302
	Recall	0.9574	0.0369	0.9702	0.0190	0.6044	0.4527	0.9559	0.0369
	AUC	0.9495	0.0437	0.9911	0.0056	0.6080	0.4802	0.9838	0.0141
	F1-Score	0.9478	0.0250	0.9721	0.0136	0.6127	0.4506	0.9570	0.0222
SM	Accuracy	0.9328	0.0339	0.9635	0.0146	0.5526	0.4623	0.9453	0.0243
	Precision	0.9471	0.0412	0.9788	0.0214	0.5802	0.4516	0.9665	0.0299
	Recall	0.9525	0.0294	0.9656	0.0166	0.5571	0.4587	0.9504	0.0327
	AUC	0.9512	0.0453	0.9907	0.0044	0.5592	0.4877	0.9836	0.0113
	F1-Score	0.9490	0.0247	0.9719	0.0110	0.5666	0.4569	0.9577	0.0191

Table 11: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on the Breast Cancer dataset under the 70/30 split, $n = 180$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.9408	0.0268	0.9630	0.0157	0.5881	0.4535	0.9482	0.0282
	Precision	0.9476	0.0358	0.9729	0.0221	0.6122	0.4449	0.9619	0.0324
	Recall	0.9642	0.0206	0.9709	0.0133	0.5926	0.4533	0.9600	0.0301
	AUC	0.9625	0.0362	0.9915	0.0048	0.5968	0.4812	0.9836	0.0155
	F1-Score	0.9553	0.0194	0.9717	0.0118	0.6005	0.4508	0.9603	0.0216
IS	Accuracy	0.9370	0.0360	0.9632	0.0180	0.6883	0.4233	0.9478	0.0281
	Precision	0.9450	0.0400	0.9720	0.0239	0.7055	0.4196	0.9616	0.0302
	Recall	0.9611	0.0326	0.9723	0.0156	0.6927	0.4261	0.9594	0.0330
	AUC	0.9569	0.0472	0.9910	0.0049	0.7033	0.4454	0.9828	0.0173
	F1-Score	0.9523	0.0273	0.9719	0.0134	0.6978	0.4239	0.9599	0.0218
MRSSU	Accuracy	0.9313	0.0399	0.9604	0.0147	0.6163	0.4473	0.9441	0.0241
	Precision	0.9395	0.0460	0.9709	0.0218	0.6383	0.4388	0.9584	0.0305
	Recall	0.9592	0.0368	0.9690	0.0132	0.6210	0.4482	0.9571	0.0291
	AUC	0.9554	0.0532	0.9912	0.0045	0.6253	0.4743	0.9842	0.0138
	F1-Score	0.9482	0.0300	0.9697	0.0109	0.6407	0.4404	0.9572	0.0181
SMOTE	Accuracy	0.9377	0.0342	0.9618	0.0198	0.5950	0.4526	0.9425	0.0269
	Precision	0.9450	0.0404	0.9714	0.0262	0.6163	0.4436	0.9567	0.0341
	Recall	0.9624	0.0236	0.9708	0.0158	0.6024	0.4543	0.9567	0.0291
	AUC	0.9556	0.0540	0.9910	0.0058	0.6054	0.4791	0.9823	0.0188
	F1-Score	0.9531	0.0247	0.9708	0.0146	0.6137	0.4484	0.9560	0.0201
SRS	Accuracy	0.9425	0.0279	0.9607	0.0176	0.5711	0.4586	0.9432	0.0318
	Precision	0.9504	0.0371	0.9726	0.0220	0.5950	0.4459	0.9618	0.0300
	Recall	0.9642	0.0234	0.9676	0.0162	0.5785	0.4583	0.9520	0.0413
	AUC	0.9656	0.0427	0.9903	0.0051	0.5782	0.4846	0.9814	0.0200
	F1-Score	0.9566	0.0198	0.9699	0.0132	0.5848	0.4538	0.9560	0.0258
Stratified	Accuracy	0.9382	0.0315	0.9662	0.0172	0.5867	0.4557	0.9474	0.0346
	Precision	0.9463	0.0392	0.9753	0.0256	0.6064	0.4475	0.9617	0.0320
	Recall	0.9617	0.0240	0.9737	0.0140	0.5935	0.4583	0.9583	0.0369
	AUC	0.9575	0.0480	0.9924	0.0052	0.5957	0.4810	0.9813	0.0274
	F1-Score	0.9533	0.0228	0.9742	0.0126	0.5982	0.4543	0.9595	0.0275
SM	Accuracy	0.9329	0.0350	0.9596	0.0191	0.5395	0.4579	0.9445	0.0241
	Precision	0.9421	0.0417	0.9710	0.0253	0.5656	0.4488	0.9598	0.0278
	Recall	0.9580	0.0299	0.9677	0.0149	0.5453	0.4583	0.9559	0.0270
	AUC	0.9509	0.0513	0.9896	0.0054	0.5467	0.4873	0.9838	0.0117
	F1-Score	0.9493	0.0259	0.9691	0.0142	0.5533	0.4552	0.9574	0.0185

Table 12: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on Breast Cancer dataset using 5-fold cross-validation, $n = 49$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.9145	0.0291	0.9572	0.0101	0.6099	0.2739	0.9284	0.0172
	Precision	0.9211	0.0340	0.9682	0.0182	0.6342	0.2662	0.9518	0.0222
	Recall	0.9534	0.0185	0.9671	0.0099	0.6130	0.2751	0.9400	0.0270
	AUC	0.9367	0.0409	0.9903	0.0015	0.6168	0.2910	0.9721	0.0140
	F1-Score	0.9366	0.0205	0.9675	0.0071	0.6230	0.2711	0.9454	0.0132
IS	Accuracy	0.9150	0.0270	0.9565	0.0111	0.5798	0.2859	0.9267	0.0174
	Precision	0.9236	0.0302	0.9671	0.0200	0.6048	0.2788	0.9509	0.0231
	Recall	0.9509	0.0221	0.9672	0.0090	0.5829	0.2868	0.9383	0.0287
	AUC	0.9388	0.0353	0.9899	0.0020	0.5855	0.3032	0.9705	0.0164
	F1-Score	0.9367	0.0193	0.9670	0.0077	0.5931	0.2835	0.9440	0.0136
MRSSU	Accuracy	0.9126	0.0280	0.9573	0.0092	0.5993	0.2661	0.9271	0.0181
	Precision	0.9195	0.0292	0.9679	0.0163	0.6237	0.2601	0.9501	0.0208
	Recall	0.9516	0.0217	0.9675	0.0090	0.6012	0.2666	0.9396	0.0242
	AUC	0.9319	0.0397	0.9899	0.0020	0.6048	0.2818	0.9682	0.0155
	F1-Score	0.9350	0.0200	0.9675	0.0065	0.6120	0.2637	0.9445	0.0139
SMOTE	Accuracy	0.9125	0.0300	0.9573	0.0101	0.6060	0.2725	0.9270	0.0190
	Precision	0.9194	0.0326	0.9679	0.0174	0.6309	0.2650	0.9487	0.0247
	Recall	0.9517	0.0196	0.9675	0.0084	0.6087	0.2724	0.9412	0.0252
	AUC	0.9341	0.0437	0.9901	0.0017	0.6131	0.2897	0.9701	0.0141
	F1-Score	0.9349	0.0214	0.9676	0.0071	0.6192	0.2693	0.9445	0.0143
SRS	Accuracy	0.9154	0.0311	0.9563	0.0100	0.6375	0.2320	0.9266	0.0182
	Precision	0.9229	0.0320	0.9666	0.0185	0.6617	0.2265	0.9494	0.0234
	Recall	0.9521	0.0234	0.9675	0.0093	0.6394	0.2326	0.9399	0.0266
	AUC	0.9371	0.0420	0.9899	0.0023	0.6469	0.2455	0.9702	0.0140
	F1-Score	0.9369	0.0225	0.9669	0.0070	0.6501	0.2298	0.9442	0.0139
Stratified	Accuracy	0.9163	0.0240	0.9566	0.0097	0.5804	0.2619	0.9270	0.0156
	Precision	0.9251	0.0273	0.9684	0.0176	0.6064	0.2542	0.9527	0.0224
	Recall	0.9510	0.0215	0.9661	0.0101	0.5841	0.2621	0.9368	0.0250
	AUC	0.9405	0.0335	0.9902	0.0017	0.5864	0.2787	0.9697	0.0146
	F1-Score	0.9375	0.0175	0.9671	0.0069	0.5948	0.2585	0.9443	0.0119
SM	Accuracy	0.9191	0.0245	0.9578	0.0093	0.5999	0.2618	0.9292	0.0154
	Precision	0.9259	0.0280	0.9692	0.0174	0.6238	0.2561	0.9506	0.0216
	Recall	0.9546	0.0177	0.9669	0.0090	0.6015	0.2634	0.9425	0.0231
	AUC	0.9402	0.0356	0.9902	0.0017	0.6073	0.2784	0.9721	0.0130
	F1-Score	0.9397	0.0174	0.9679	0.0066	0.6121	0.2603	0.9462	0.0116

Table 13: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on Breast Cancer dataset using 5-fold cross-validation, $n = 100$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.9277	0.0243	0.9621	0.0098	0.5940	0.2809	0.9429	0.0165
	Precision	0.9369	0.0306	0.9744	0.0157	0.6174	0.2747	0.9599	0.0216
	Recall	0.9555	0.0156	0.9679	0.0067	0.5970	0.2833	0.9536	0.0172
	AUC	0.9446	0.0313	0.9905	0.0017	0.6008	0.2976	0.9821	0.0099
	F1-Score	0.9458	0.0172	0.9710	0.0071	0.6067	0.2795	0.9565	0.0121
IS	Accuracy	0.9267	0.0247	0.9608	0.0087	0.5998	0.2678	0.9398	0.0170
	Precision	0.9361	0.0310	0.9718	0.0151	0.6235	0.2601	0.9567	0.0212
	Recall	0.9548	0.0159	0.9686	0.0069	0.6032	0.2705	0.9521	0.0180
	AUC	0.9420	0.0339	0.9906	0.0021	0.6074	0.2849	0.9820	0.0093
	F1-Score	0.9450	0.0176	0.9701	0.0062	0.6128	0.2657	0.9542	0.0126
MRSSU	Accuracy	0.9282	0.0234	0.9605	0.0105	0.6141	0.2737	0.9406	0.0177
	Precision	0.9375	0.0301	0.9712	0.0175	0.6374	0.2655	0.9559	0.0218
	Recall	0.9556	0.0156	0.9687	0.0073	0.6176	0.2750	0.9542	0.0161
	AUC	0.9446	0.0328	0.9903	0.0020	0.6227	0.2914	0.9814	0.0091
	F1-Score	0.9461	0.0167	0.9699	0.0076	0.6270	0.2707	0.9549	0.0129
SMOTE	Accuracy	0.9309	0.0229	0.9612	0.0103	0.5937	0.2764	0.9401	0.0193
	Precision	0.9388	0.0297	0.9724	0.0170	0.6170	0.2684	0.9553	0.0253
	Recall	0.9587	0.0139	0.9686	0.0064	0.5980	0.2799	0.9545	0.0157
	AUC	0.9480	0.0331	0.9901	0.0019	0.6015	0.2948	0.9806	0.0115
	F1-Score	0.9483	0.0161	0.9704	0.0075	0.6069	0.2746	0.9546	0.0137
SRS	Accuracy	0.9292	0.0249	0.9611	0.0096	0.6207	0.2627	0.9417	0.0185
	Precision	0.9390	0.0310	0.9727	0.0160	0.6438	0.2552	0.9578	0.0223
	Recall	0.9557	0.0168	0.9682	0.0070	0.6244	0.2653	0.9541	0.0190
	AUC	0.9452	0.0354	0.9905	0.0020	0.6284	0.2795	0.9824	0.0093
	F1-Score	0.9469	0.0180	0.9703	0.0069	0.6336	0.2607	0.9557	0.0136
Stratified	Accuracy	0.9273	0.0250	0.9615	0.0086	0.6166	0.2566	0.9420	0.0154
	Precision	0.9368	0.0284	0.9727	0.0141	0.6397	0.2510	0.9596	0.0172
	Recall	0.9545	0.0212	0.9687	0.0054	0.6190	0.2587	0.9524	0.0185
	AUC	0.9452	0.0337	0.9904	0.0019	0.6247	0.2720	0.9817	0.0097
	F1-Score	0.9453	0.0190	0.9706	0.0062	0.6289	0.2552	0.9558	0.0116
SM	Accuracy	0.9279	0.0234	0.9606	0.0100	0.6324	0.2871	0.9388	0.0190
	Precision	0.9368	0.0305	0.9728	0.0164	0.6557	0.2803	0.9574	0.0220
	Recall	0.9561	0.01526	0.9673	0.0072	0.6351	0.2887	0.9501	0.0216
	AUC	0.9432	0.0338	0.9903	0.0016	0.6416	0.3039	0.9814	0.0081
	F1-Score	0.9460	0.0164	0.9699	0.0072	0.6449	0.2850	0.9535	0.0142

Table 14: Performance metrics (Accuracy, Precision, Recall, F1-Score, AUC) for different sampling methods on Breast Cancer dataset using 5-fold cross-validation, $n = 180$

Method	Metric	GLM		RF		SVM		XGB	
		mean	sd	mean	sd	mean	sd	mean	sd
ADASYN	Accuracy	0.9375	0.0267	0.9620	0.0115	0.5919	0.2845	0.9468	0.0193
	Precision	0.9461	0.0343	0.9729	0.0181	0.6139	0.2772	0.9619	0.0219
	Recall	0.9610	0.0137	0.9695	0.0050	0.5977	0.2916	0.9577	0.0172
	AUC	0.9561	0.0376	0.9906	0.0021	0.6014	0.3052	0.9844	0.0080
	F1-Score	0.9531	0.0190	0.9711	0.0083	0.6052	0.2846	0.9596	0.0143
IS	Accuracy	0.9386	0.0261	0.9615	0.0132	0.6286	0.2781	0.9470	0.0169
	Precision	0.9471	0.0336	0.9730	0.0199	0.6507	0.2701	0.9602	0.0214
	Recall	0.9618	0.0119	0.9685	0.0053	0.6342	0.2820	0.9597	0.0132
	AUC	0.9574	0.0365	0.9907	0.0022	0.6391	0.2974	0.9837	0.0085
	F1-Score	0.9540	0.0183	0.9706	0.0096	0.6420	0.2762	0.9598	0.0123
MRSSU	Accuracy	0.9339	0.0336	0.9614	0.0131	0.6440	0.2831	0.9451	0.0213
	Precision	0.9438	0.0401	0.9732	0.0198	0.6655	0.2750	0.9608	0.0227
	Recall	0.9568	0.0214	0.9682	0.0058	0.6500	0.2891	0.9558	0.0193
	AUC	0.9538	0.0423	0.9903	0.0025	0.6565	0.3024	0.9820	0.0127
	F1-Score	0.9499	0.0270	0.9706	0.0095	0.6571	0.2824	0.9581	0.0159
SMOTE	Accuracy	0.9376	0.0244	0.9617	0.0123	0.6255	0.2695	0.9465	0.0200
	Precision	0.9466	0.0317	0.9727	0.0179	0.6468	0.2624	0.9613	0.0242
	Recall	0.9605	0.0154	0.9691	0.0043	0.6318	0.2765	0.9578	0.0150
	AUC	0.9596	0.0301	0.9903	0.0022	0.6368	0.2886	0.9831	0.0094
	F1-Score	0.9532	0.0173	0.9708	0.0089	0.6388	0.2696	0.9594	0.0146
SRS	Accuracy	0.9355	0.0309	0.9608	0.0132	0.5772	0.2876	0.9448	0.0232
	Precision	0.9452	0.0351	0.9721	0.0200	0.5997	0.2775	0.9614	0.0249
	Recall	0.9584	0.0185	0.9684	0.0060	0.5852	0.2939	0.9552	0.0218
	AUC	0.9559	0.0394	0.9907	0.0022	0.5860	0.3087	0.9830	0.0108
	F1-Score	0.9515	0.0228	0.9702	0.0095	0.5917	0.2861	0.9580	0.0173
Stratified	Accuracy	0.9397	0.0240	0.9617	0.0121	0.6411	0.2688	0.9460	0.0197
	Precision	0.9475	0.0309	0.9729	0.0176	0.6625	0.2621	0.9609	0.0214
	Recall	0.9628	0.0120	0.9689	0.0051	0.6460	0.2741	0.9574	0.0167
	AUC	0.9585	0.0357	0.9907	0.0023	0.6526	0.2879	0.9836	0.0093
	F1-Score	0.9548	0.0167	0.9708	0.0088	0.6538	0.2682	0.9590	0.0145
SM	Accuracy	0.9387	0.0266	0.9618	0.0122	0.6407	0.2672	0.9441	0.0216
	Precision	0.9463	0.0333	0.9731	0.0183	0.6616	0.2596	0.9602	0.0240
	Recall	0.9623	0.0132	0.9689	0.0053	0.6471	0.2738	0.9552	0.0171
	AUC	0.9587	0.0371	0.9905	0.0023	0.6534	0.2863	0.9828	0.0098
	F1-Score	0.9539	0.0189	0.9709	0.0088	0.6538	0.2668	0.9575	0.0158

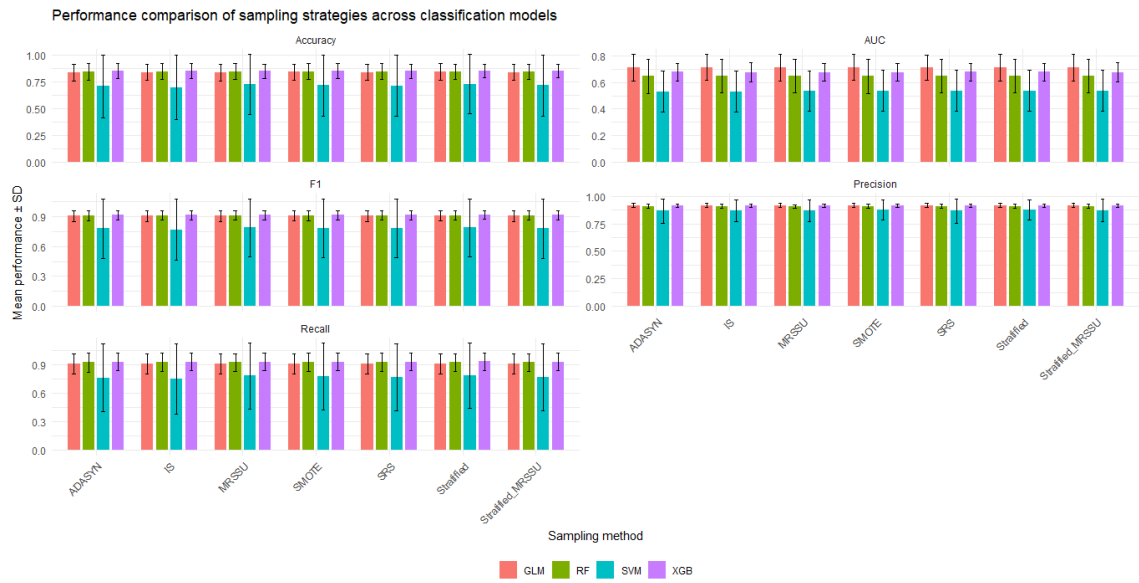


Figure 3: Comparison of sampling methods for class ratio (0.1, 0.9) and sample size 180.

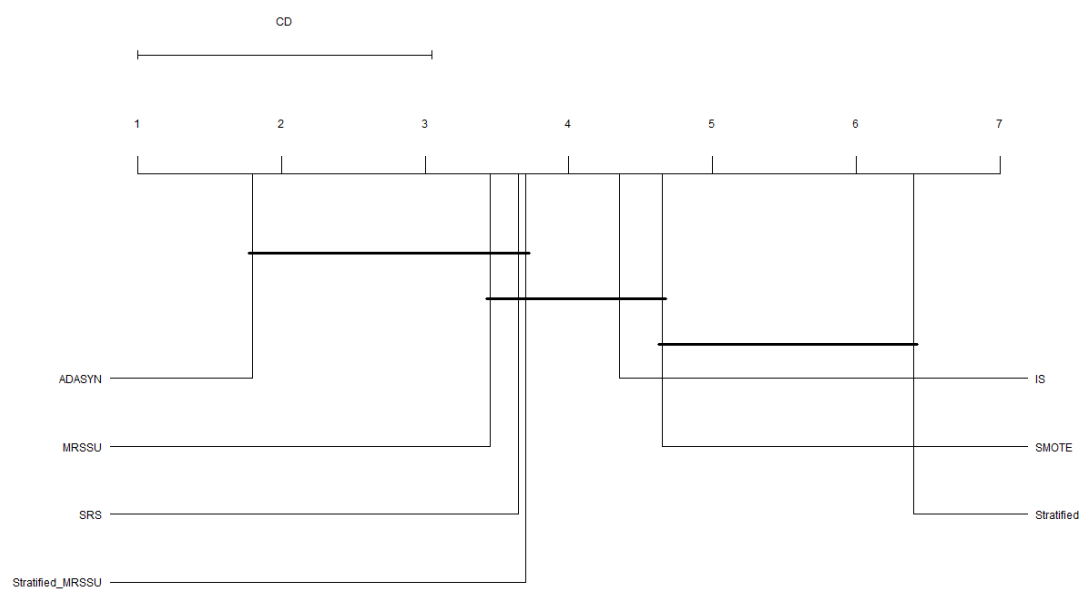


Figure 4: Critical Difference Diagram for Class Ratio (0.1, 0.9) and Sample Size 180

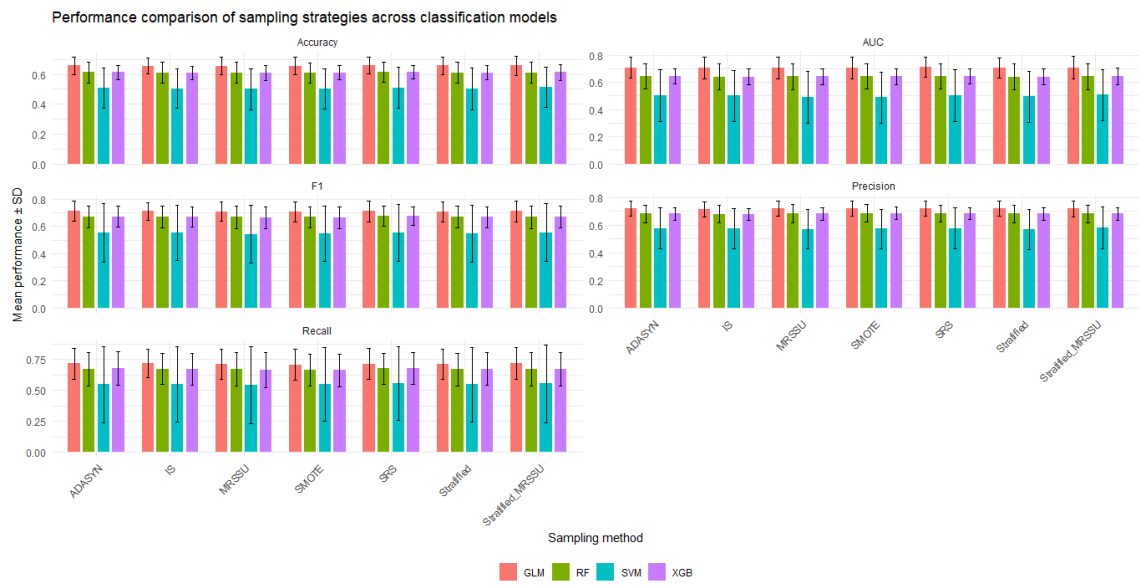


Figure 5: Comparison of sampling methods for class ratio (0.4, 0.6) and sample size 49.

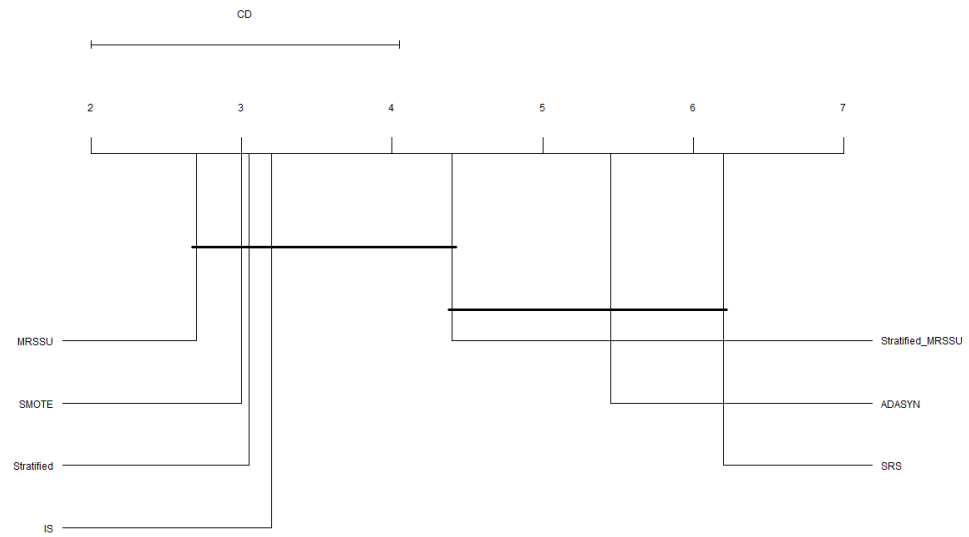


Figure 6: Critical Difference Diagram for Class Ratio (0.4, 0.6) and sample size 49.

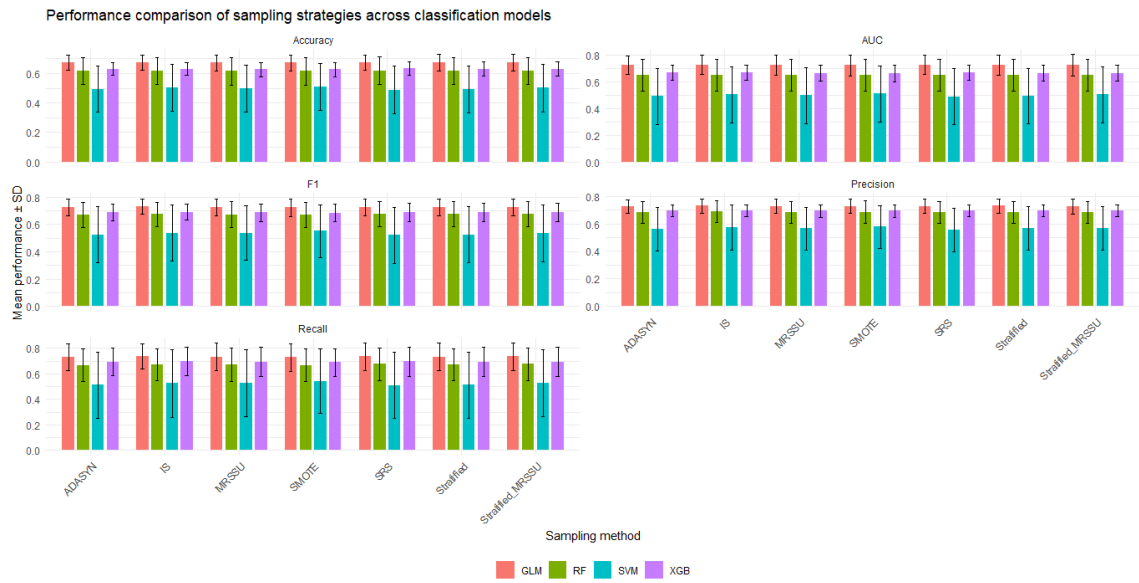


Figure 7: Comparison of sampling methods for class ratio (0.4, 0.6) and sample size 100.

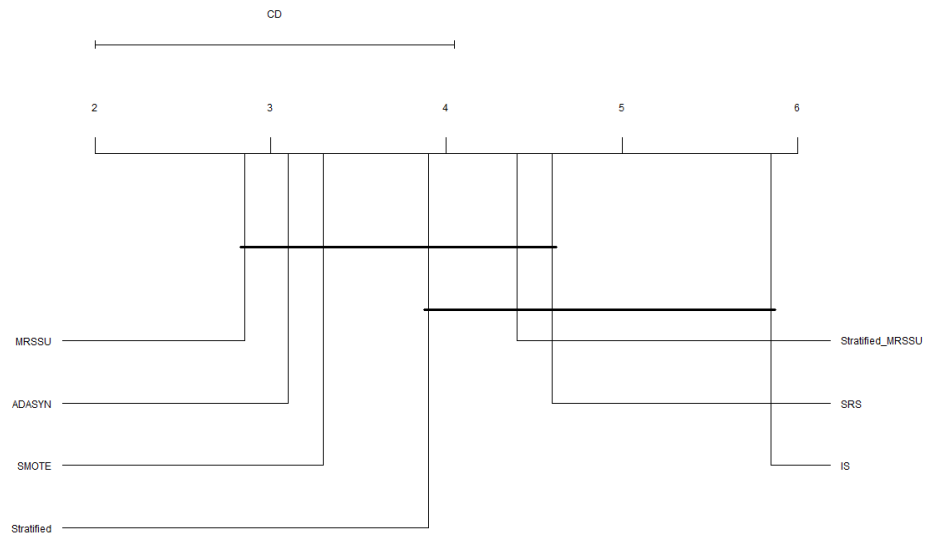


Figure 8: Critical Difference Diagram for Class Ratio (0.4, 0.6) and sample size 100.

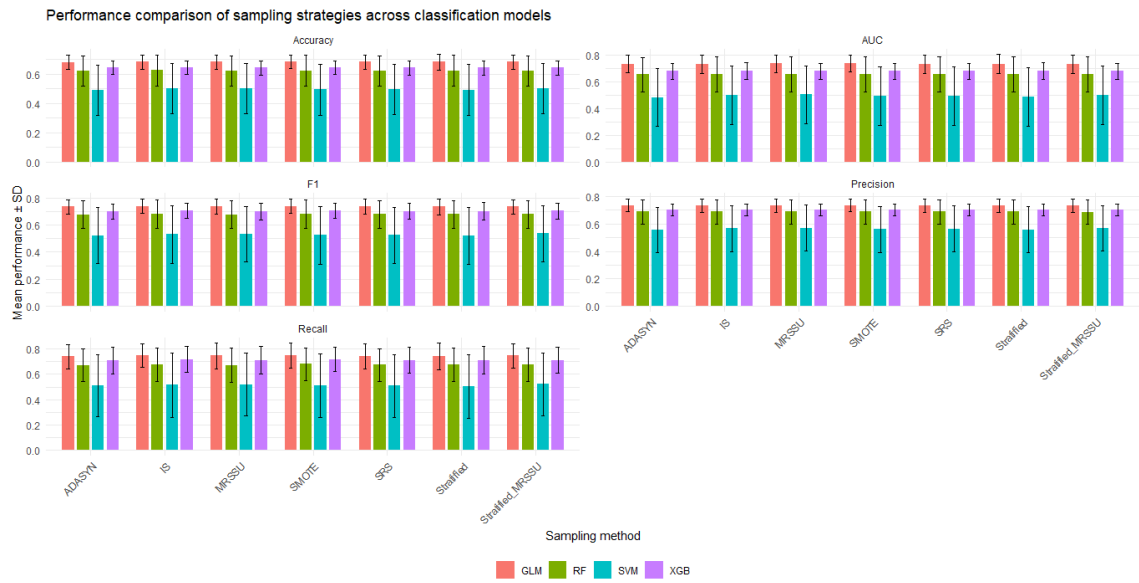


Figure 9: Comparison of sampling methods for class ratio (0.4, 0.6) and sample size 180.

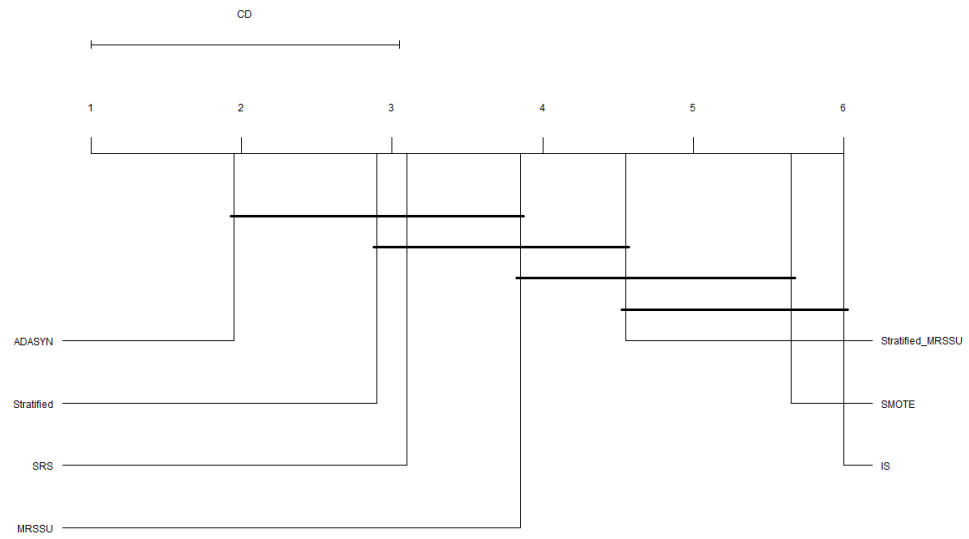


Figure 10: Critical Difference Diagram for Class Ratio (0.4, 0.6) and sample size 180.

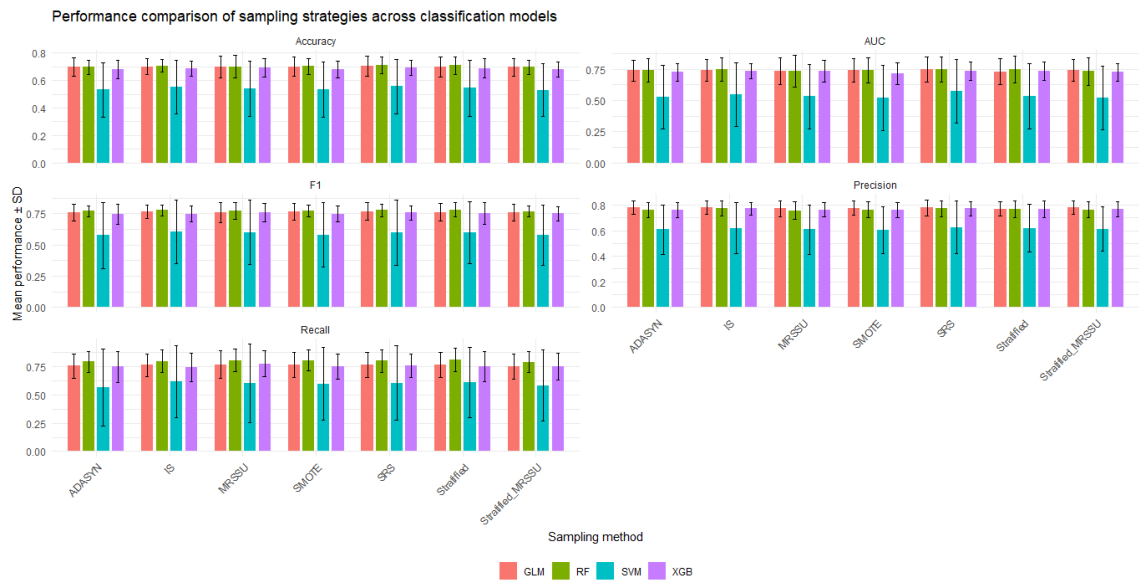


Figure 11: Comparison of Sampling Methods for the Pima Indians Diabetes dataset under the 70/30 split, $n=49$

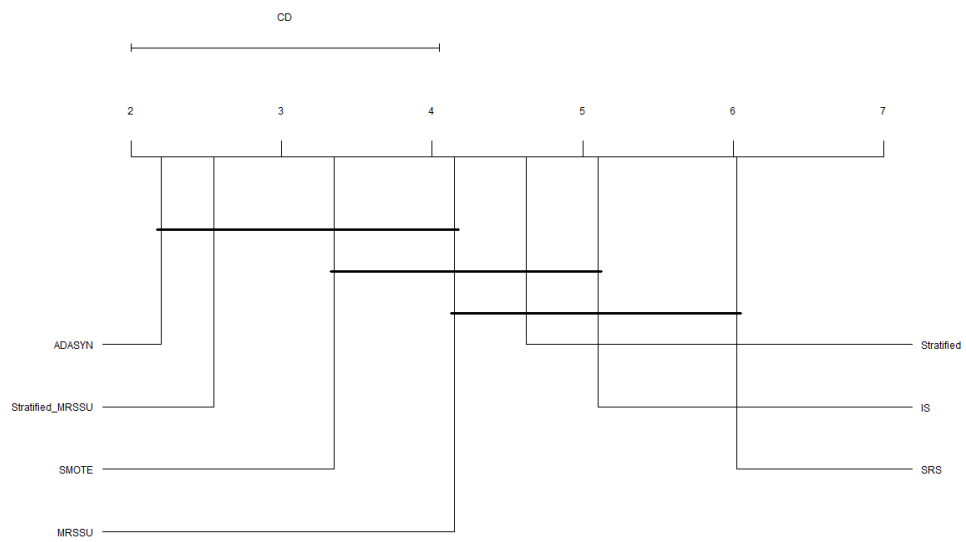


Figure 12: Critical Difference Diagram for the Pima Indians Diabetes dataset under the 70/30 split, $n=49$

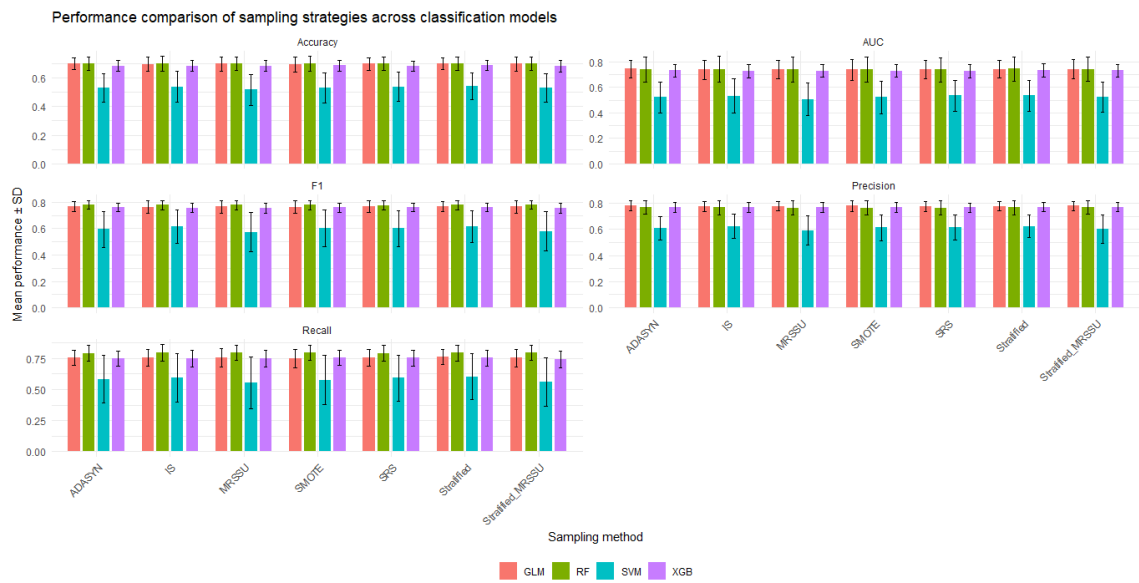


Figure 13: Comparison of Sampling Methods for the Pima Indians Diabetes dataset using 5-fold cross-validation, $n=49$

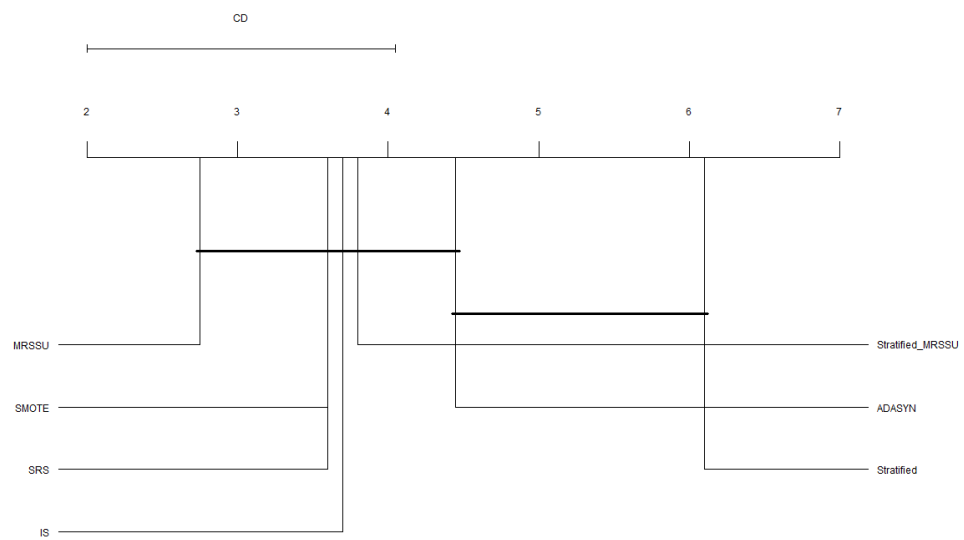


Figure 14: Critical Difference Diagram for the Pima Indians Diabetes dataset using 5-fold cross-validation, $n=49$

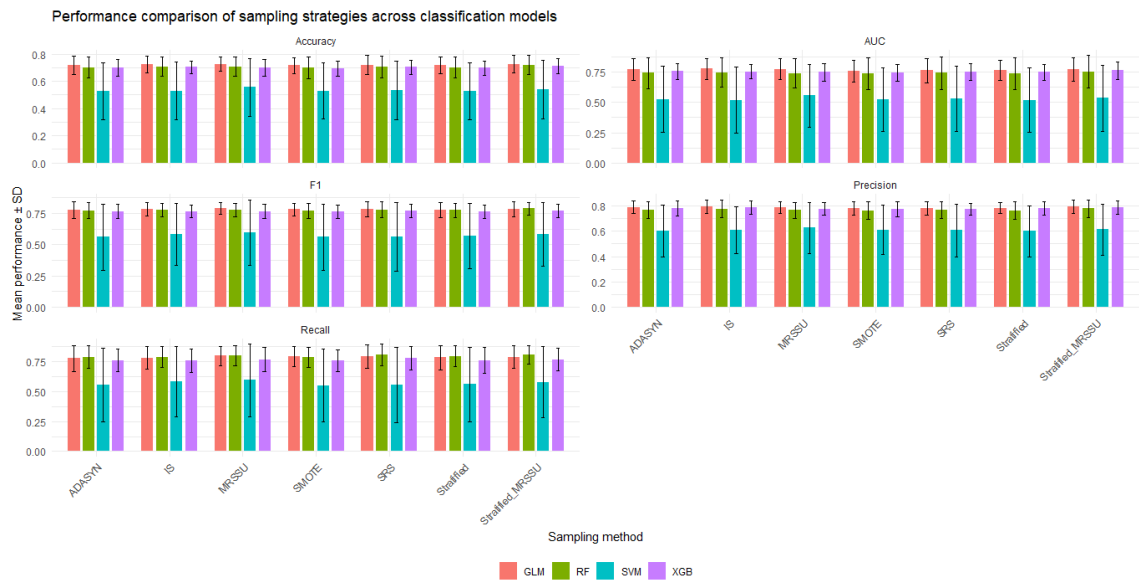


Figure 15: Comparison of Sampling Methods for the Pima Indians Diabetes dataset under the 70/30 split, $n=100$

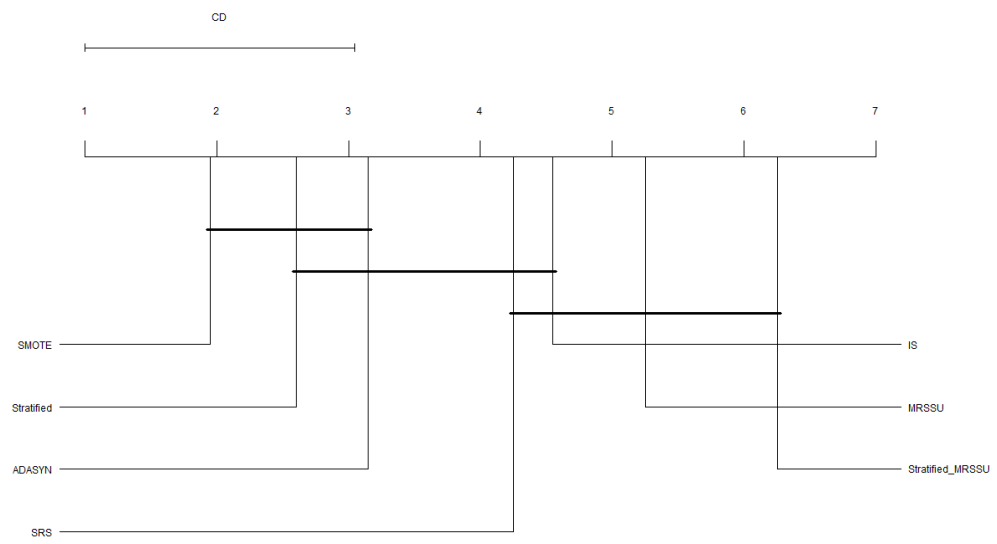


Figure 16: Critical Difference Diagram for the Pima Indians Diabetes dataset under the 70/30 split, $n=100$

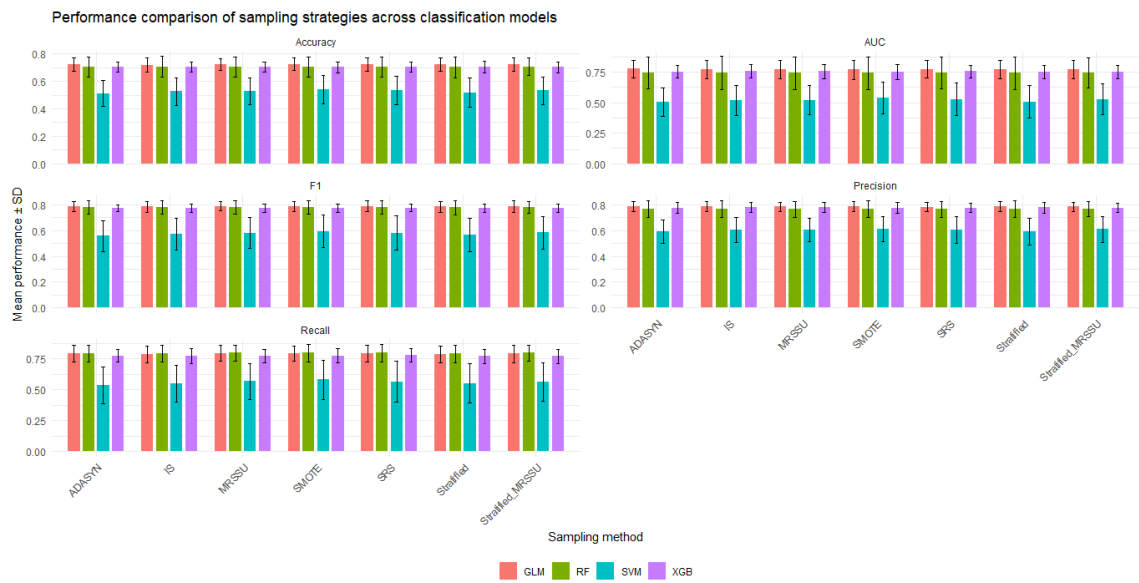


Figure 17: Comparison of Sampling Methods for the Pima Indians Diabetes dataset using 5-fold cross-validation, $n=100$

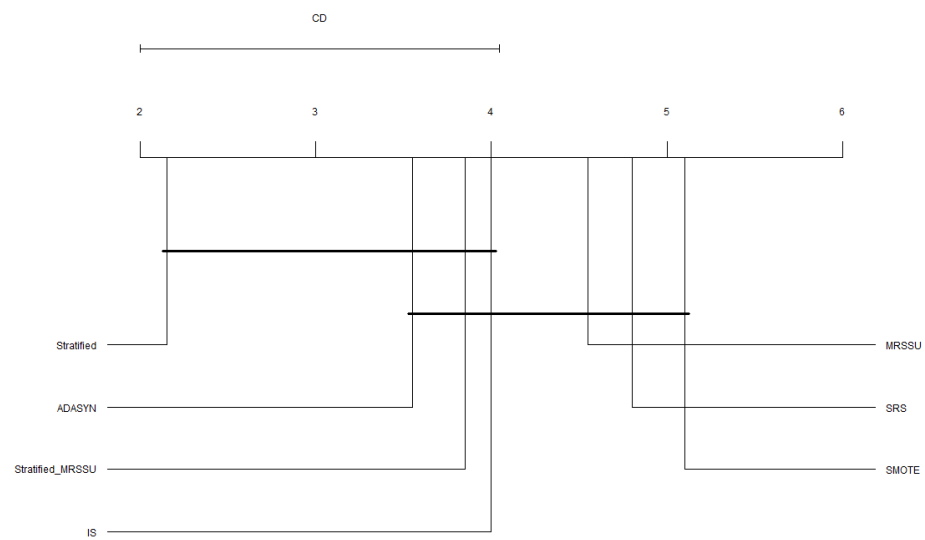


Figure 18: Critical Difference Diagram for the Pima Indians Diabetes dataset using 5-fold cross-validation, $n=100$

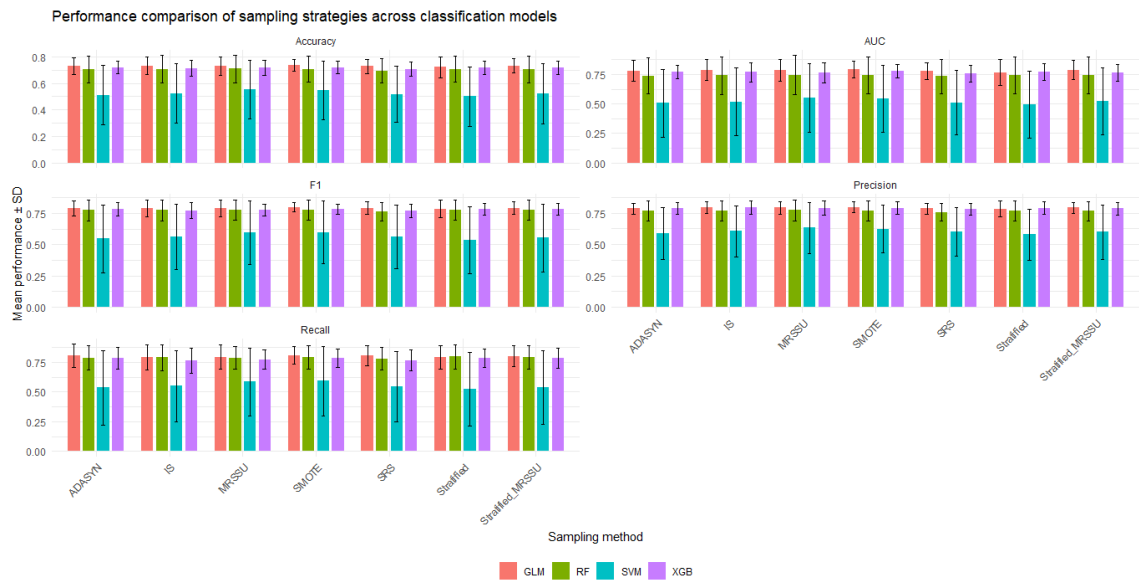


Figure 19: Comparison of Sampling Methods for the Pima Indians Diabetes dataset under the 70/30 split, n=180

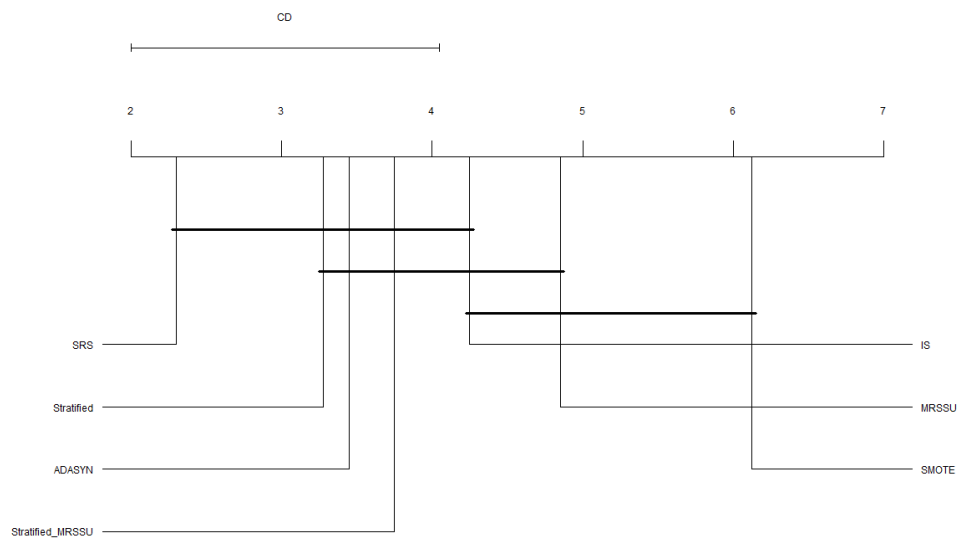


Figure 20: Critical Difference Diagram for the Pima Indians Diabetes dataset under the 70/30 split, $n=180$

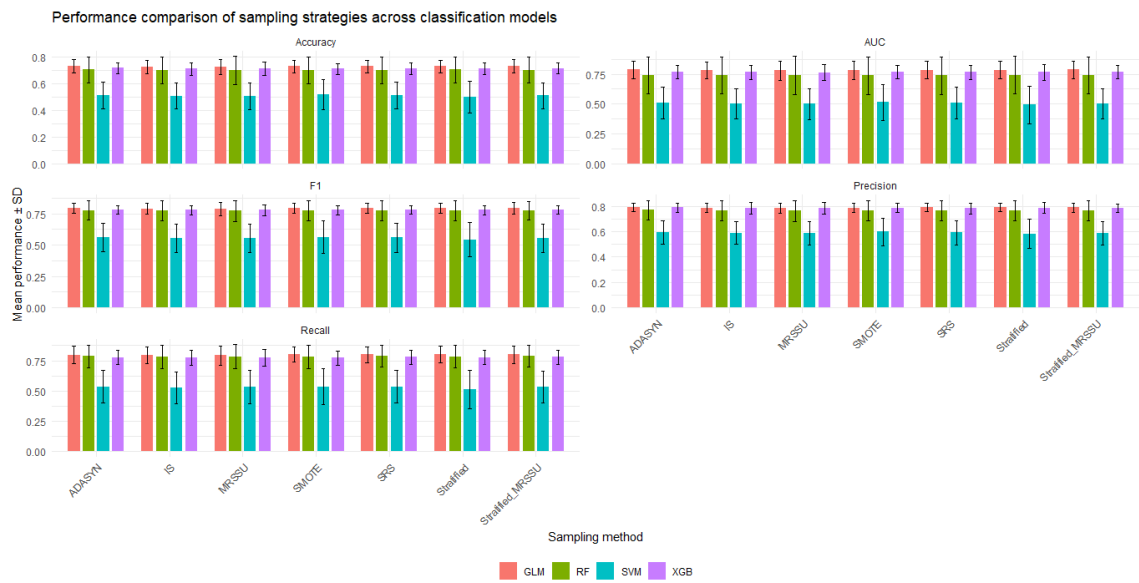


Figure 21: Comparison of Sampling Methods for the Pima Indians Diabetes dataset using 5-fold cross-validation, n=180

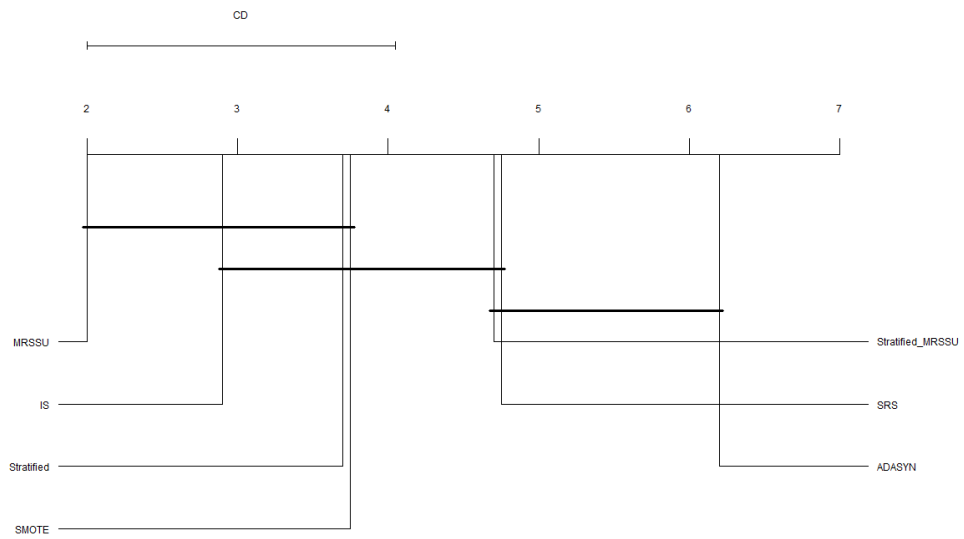


Figure 22: Critical Difference Diagram for the Pima Indians Diabetes dataset using 5-fold cross-validation, $n=180$

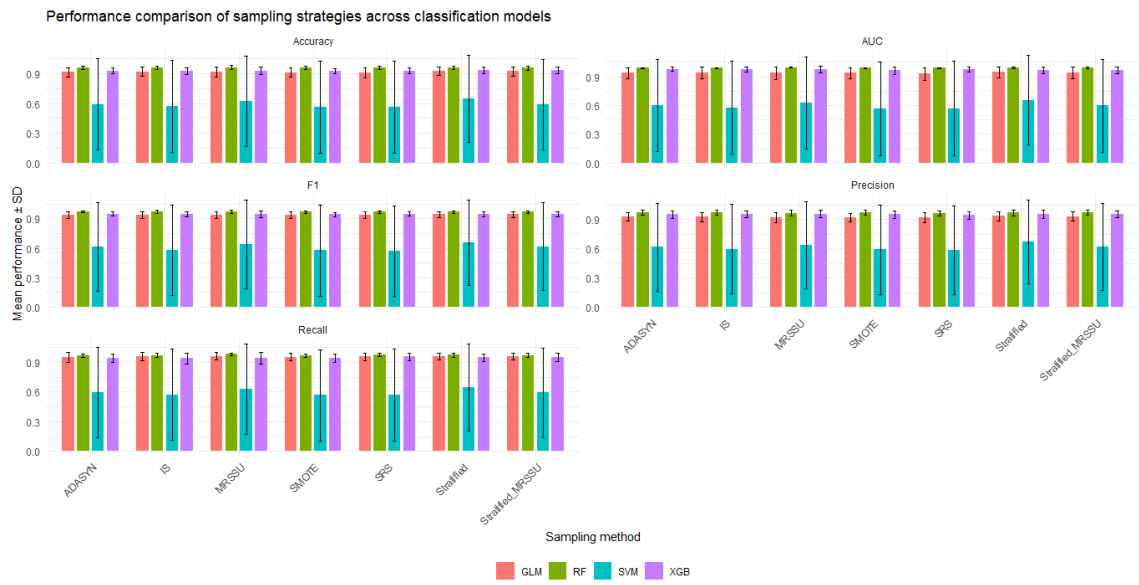


Figure 23: Comparison of Sampling Methods for Breast Cancer dataset under the 70/30 split, $n=49$

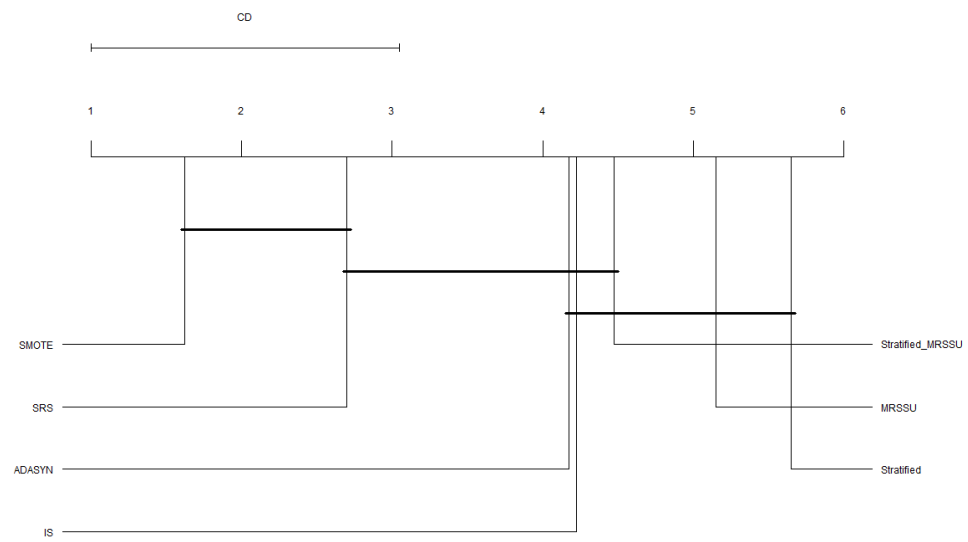


Figure 24: Critical Difference Diagram for Breast Cancer dataset under the 70/30 split, $n=49$

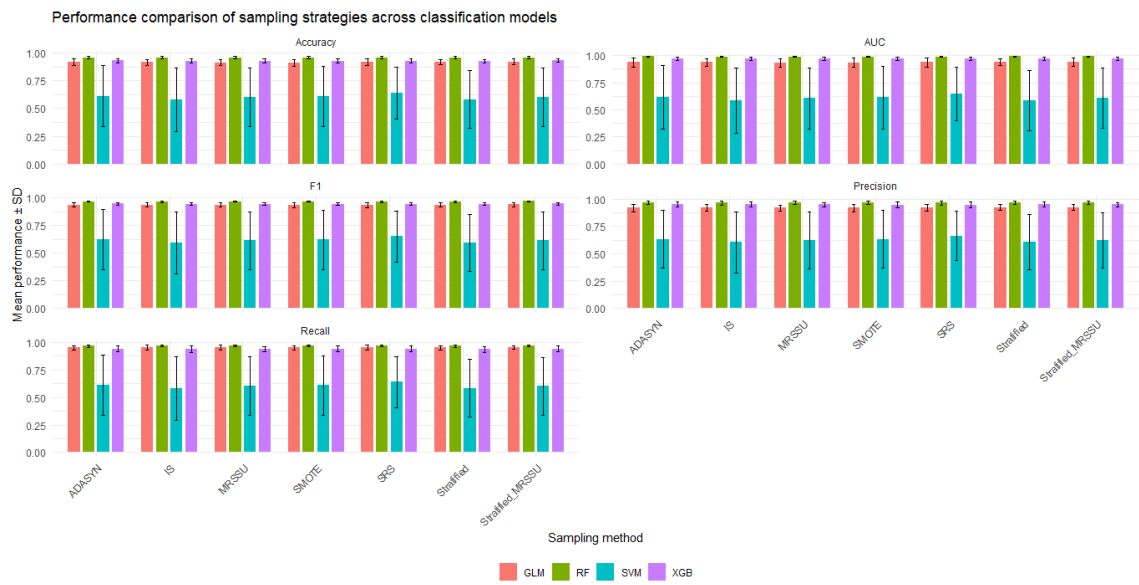


Figure 25: Comparison of Sampling Methods for Breast Cancer dataset using 5-fold cross-validation, $n=49$

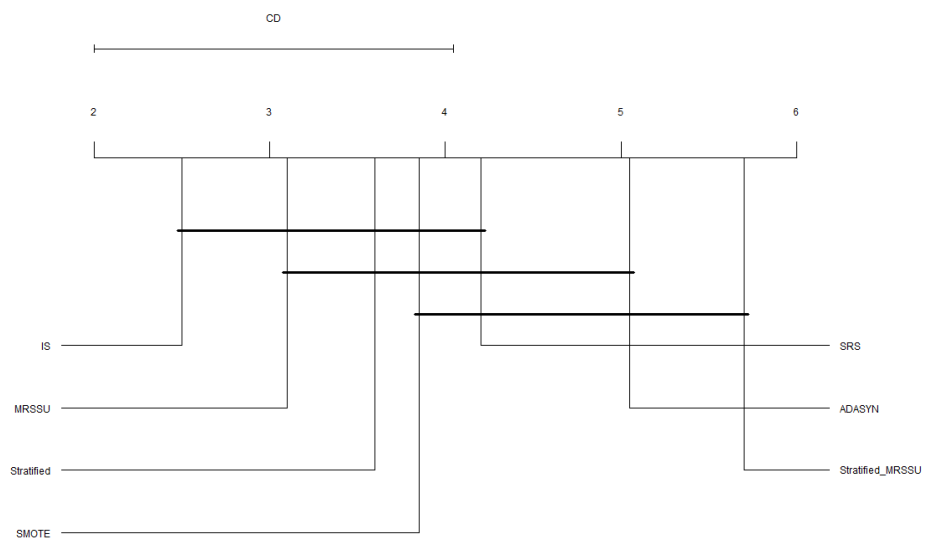


Figure 26: Critical Difference Diagram for Breast Cancer dataset using 5-fold cross-validation, $n=49$

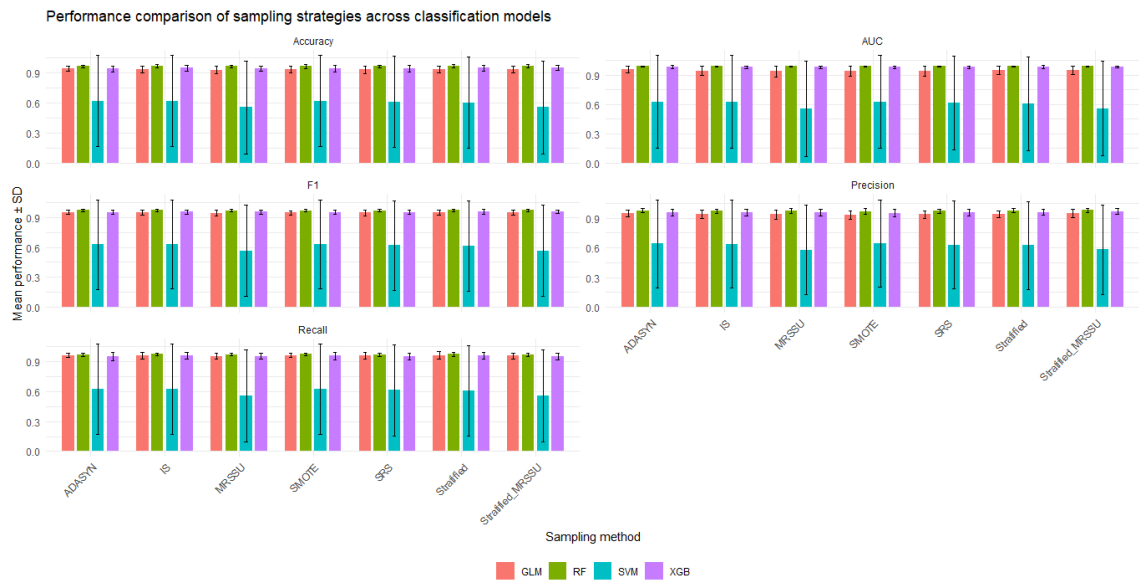


Figure 27: Comparison of Sampling Methods for Breast Cancer dataset under the 70/30 split, $n=100$

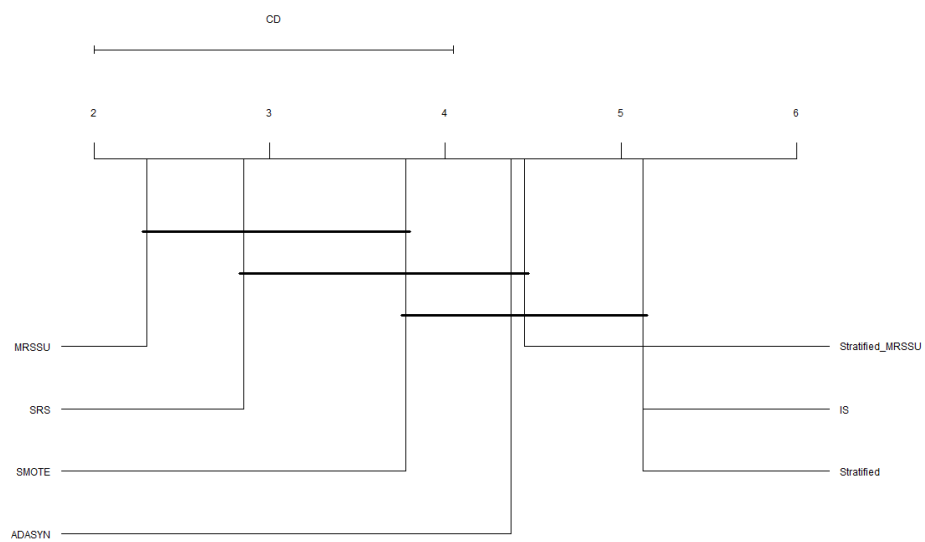


Figure 28: Critical Difference Diagram for Breast Cancer dataset under the 70/30 split, $n=100$

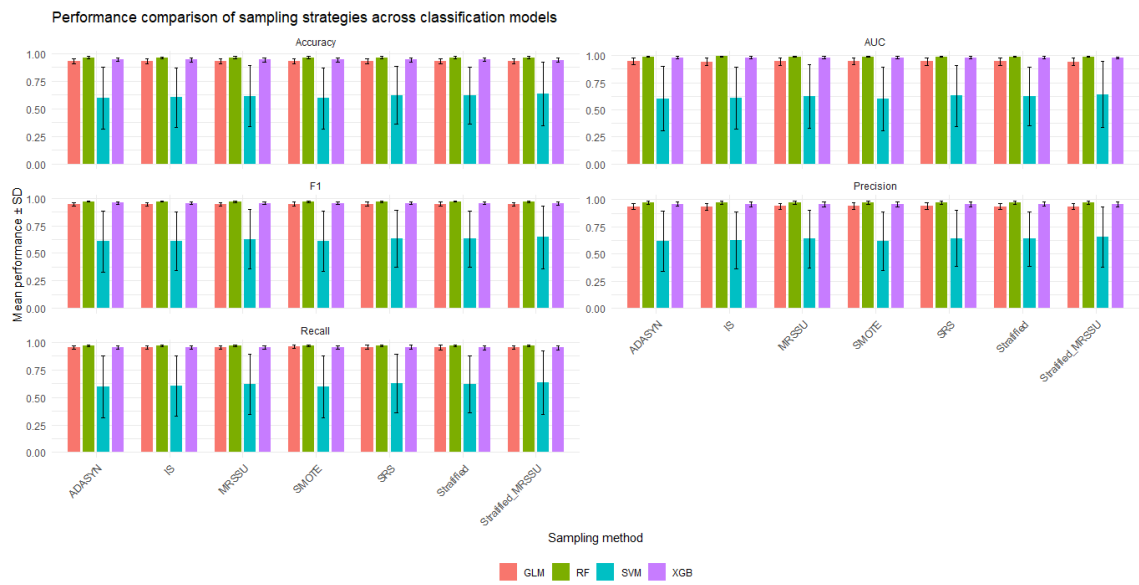


Figure 29: Comparison of Sampling Methods for Breast Cancer dataset using 5-fold cross-validation, n=100

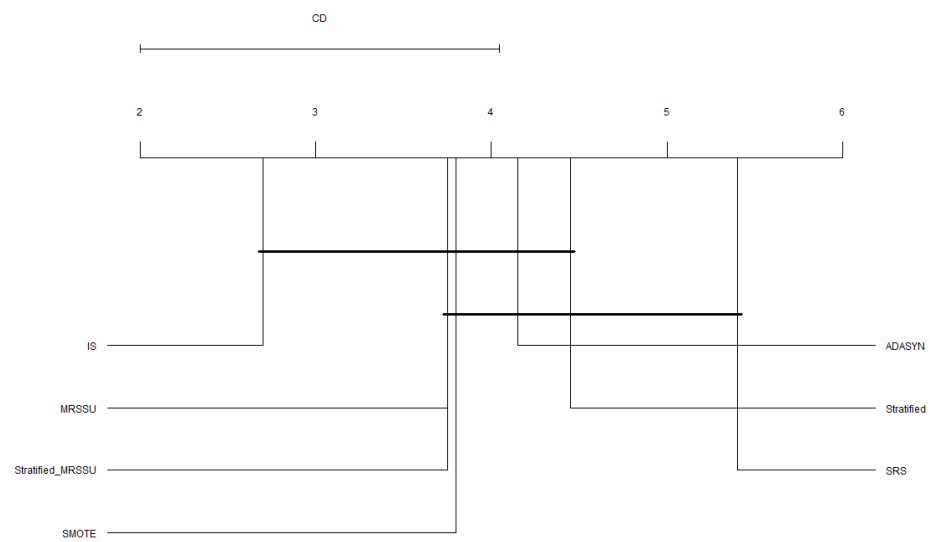


Figure 30: Critical Difference Diagram for Breast Cancer dataset using 5-fold cross-validation, n=100

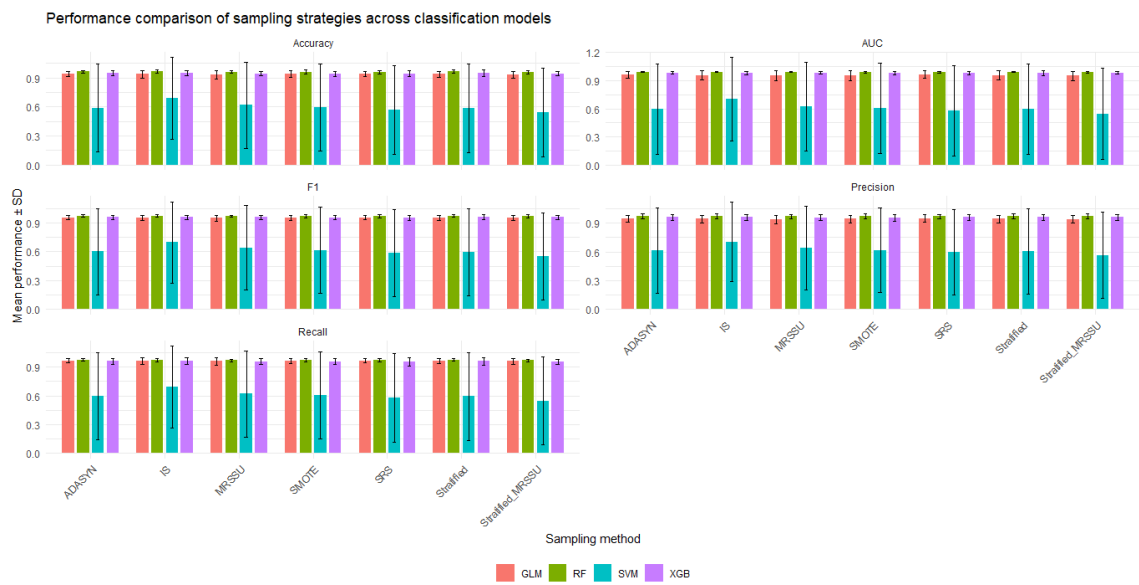


Figure 31: Comparison of Sampling Methods for Breast Cancer dataset under the 70/30 split, $n=180$

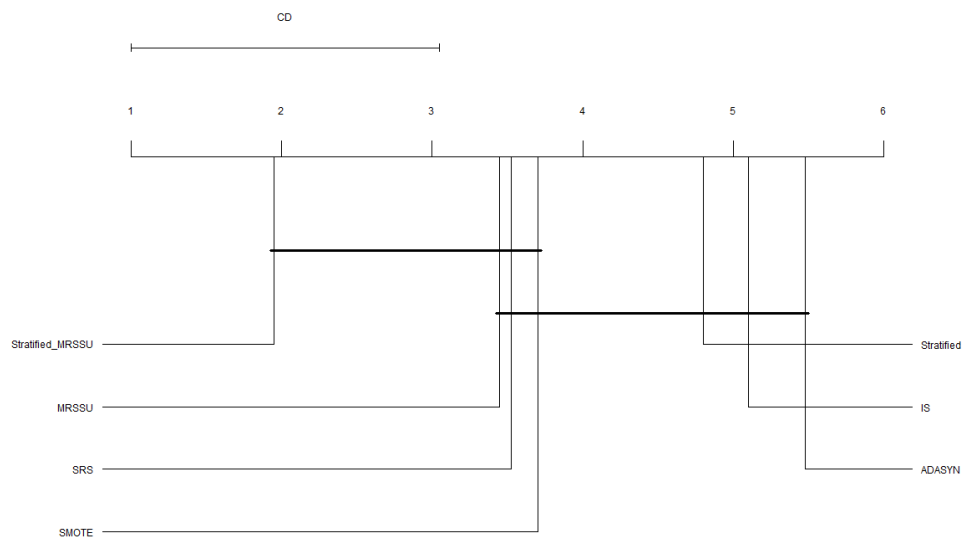


Figure 32: Critical Difference Diagram for Breast Cancer dataset under the 70/30 split, $n=180$

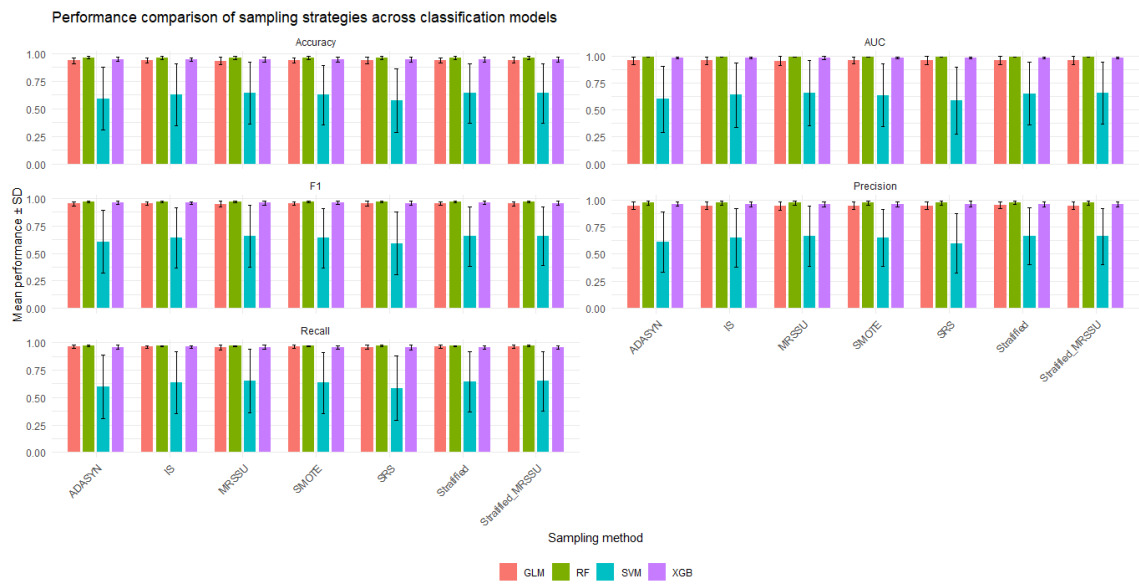


Figure 33: Comparison of Sampling Methods for Breast Cancer dataset using 5-fold cross-validation, $n=180$

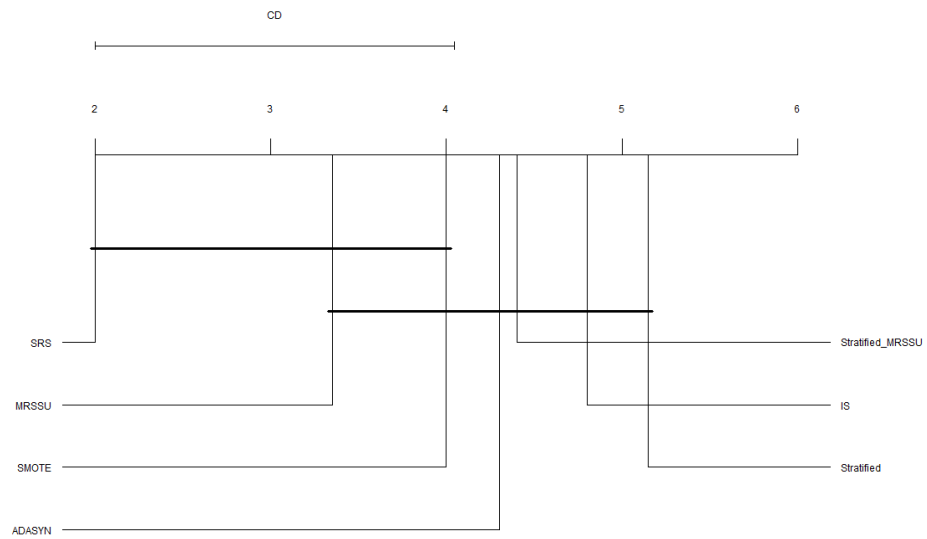


Figure 34: Critical Difference Diagram for Breast Cancer dataset using 5-fold cross-validation, $n=180$