CSC 4850 Team 5 Project Report

Max Arendale, David Okerinde, Josh Pauldin, Shrey Patel

March 25, 2024

1 Model Results

Datasets	Metrics	Project Results								DrugOrchestra Results			
		KNN	Linear Regression	Logistic Regression	Naive Bayes	Decision Tree	Random Forest	Gradient Boosting	SVM	SVM	Random Forest	STL-NN	DrugOrchestra
STITCH	AUROC	0.659±0.066	0.775±0.05	0.763±0.053	0.73±0.069	0.573±0.047	0.752±0.037	0.736±0.039	0.74±0.013	0.668±0.048	0.666±0.032	0.827±0.031	**0.833±0.047**
	AUPRC	0.214±0.056	0.34±0.132	0.334±0.108	0.303±0.119	0.146±0.028	0.295±0.087	0.311±0.087	0.295±0.059	0.320±0.092	0.349±0.053	0.583±0.034	**0.587±0.088**
Drugbank	AUROC	0.688±0.024	0.689±0.062	0.665±0.078	0.716±0.103	0.583±0.058	0.727±0.076	0.652±0.057	**0.784±0.049**	0.625±0.042	0.616±0.035	0.748±0.034	0.752±0.023
	AUPRC	0.368±0.119	0.503±0.094	0.448±0.102	0.491±0.126	0.114±0.041	0.392±0.117	0.296±0.118	0.48±0.125	0.260±0.098	0.269±0.055	0.530±0.041	**0.540±0.012**
Repur	AUROC	0.602±0.041	0.669±0.009	0.672±0.014	0.635±0.053	0.579±0.054	0.644±0.043	0.604±0.038	0.66±0.042	0.530±0.010	0.585±0.049	0.691±0.014	**0.703±0.031**
	AUPRC	0.138±0.018	0.196±0.016	0.234±0.055	0.222±0.029	0.143±0.035	0.218±0.04	0.214±0.032	0.216±0.033	0.125±0.018	0.221±0.066	0.293±0.023	**0.321±0.036**
PDX	SCC	0.244±0.073	0.231±0.083	N/A	N/A	0.26±0.062	0.284±0.044	0.276±0.054	0.266±0.027	**0.491±0.015**	0.436±0.017	0.468±0.094	0.465±0.014
	MSE	1.049±0.138	1.152±0.072	N/A	N/A	1.527±0.162	1.006±0.058	0.958±0.131	0.989±0.164	**0.824±0.034**	0.846±0.034	0.856±0.028	0.840±0.023
GDSC	SCC	0.058±0.04	0.061±0.036	N/A	N/A	0.043±0.033	0.069±0.021	0.053±0.038	0.091±0.011	0.235±0.039	0.151±0.020	0.394±0.022	**0.410±0.010**
	MSE	1.761±0.37	1.441±0.182	N/A	N/A	2.736±0.206	1.47±0.163	1.324±0.209	1.31±0.2	1.150±0.086	1.110±0.048	0.930±0.024	**0.814±0.061**
CCLE	SCC	0.104±0.022	0.177±0.051	N/A	N/A	0.102±0.032	0.063±0.035	0.077±0.05	0.148±0.05	0.206±0.016	0.118±0.012	0.394±0.022	**0.410±0.010**
	MSE	1.432±0.199	1.102±0.092	N/A	N/A	2.023±0.179	1.165±0.106	1.093±0.135	1.086±0.166	1.472±0.046	1.147±0.031	0.930±0.026	**0.882±0.037**
SIDER	AUROC	0.567±0.084	0.541±0.088	0.544±0.068	**0.701±0.019**	0.562±0.019	0.69±0.022	0.672±0.023	0.672±0.026	0.591±0.026	0.640±0.026	0.660±0.026	0.671±0.036
	AUPRC	0.192±0.045	0.252±0.063	0.242±0.064	**0.359±0.008**	0.164±0.01	0.29±0.029	0.296±0.02	0.33±0.012	0.231±0.022	0.295±0.035	0.312±0.016	0.321±0.036
OFFSIDE	AUROC	0.544±0.042	0.525±0.072	0.501±0.049	0.584±0.075	0.535±0.021	0.603±0.063	0.544±0.043	0.59±0.061	0.629±0.009	0.638±0.029	0.640±0.025	**0.686±0.011**
	AUPRC	0.328±0.031	0.369±0.058	0.343±0.037	0.419±0.09	0.311±0.012	0.389±0.049	0.372±0.043	0.383±0.056	0.399±0.005	0.435±0.013	0.432±0.022	0.457±0.037

2 Transferability

2.1 Datasets

2.1.1 KNN

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	-27.817	-16.944	-130.489	-193.516	-186.817	-9.042	-7.458
Drugbank	-20.654	0	-16.944	-101.697	**261.581**	-74.514	-12.433	-6.704
Repur	-24.127	-27.326	0	-135.984	84.426	**5.691**	-11.817	-7.58
PDX	-23.608	-26.835	-16.382	0	138.82	-164.631	-27.738	-30.357
GDSC	-24.127	-31.11	-16.382	-106.653	0	-27.786	-31.101	-23.436
CCLE	**-20.298**	**-20.855**	**-15.755**	**-22.845**	-86.559	0	-0.997	-11.732
SIDER	-26.89	-29.29	-22.555	-70.148	-54.475	-58.301	0	**-5.32**
OFFSIDE	-25.851	-30.763	-22.555	-81.109	10.041	-213.457	**8.225**	0

2.1.2 Linear Regression

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	-10.76	-34.767	-107.559	-431.047	**-13.139**	9.93	-6.069
Drugbank	-19.762	0	-38.154	-92.353	68.626	-39.538	-37.837	8.854
Repur	-30.646	-28.354	0	-104.651	-175.229	-67.986	-19.095	4.533
PDX	-20.692	**-6.837**	**-15.102**	0	**131.731**	-120.997	-17.038	-3.128
GDSC	-42.787	-43.179	-27.579	**-42.079**	0	-74.999	7.815	-8.465
CCLE	-43.298	-32.334	-16.944	-102.214	-7.407	0	12.692	**9.471**
SIDER	**-6.97**	-10.414	-17.122	-103.511	-145.008	-99.709	0	6.421
OFFSIDE	-45.392	-44.506	-31.737	-128.227	-66.784	-40.003	**25.324**	0

2.1.3 Decision Tree

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	-8.749	-17.269	-98.346	-125.469	-171.94	-21.379	-2.568
Drugbank	-21.054	0	-9.471	-78.516	-159.347	-66.123	-27.283	-1.383
Repur	-5.492	-22.68	0	-92.939	**44.939**	**-20.575**	-17.934	**10.837**
PDX	-21.372	**-0.537**	-21.394	0	-71.851	-135.211	-47.311	-12.198
GDSC	-23.96	-20.372	-18.467	**-55.885**	0	-26.813	-46.058	-5.211
CCLE	**-1.007**	-12.829	-13.871	-77.354	-384.506	0	-26.198	-0.754
SIDER	-17.147	-10.913	**-7.236**	-104.795	9.405	-183.731	0	3.445
OFFSIDE	-8.556	-9.911	-12.704	-80.088	-45.674	-97.485	**-14.944**	0

2.1.4 Random Forest

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	**-21.993**	-20.805	-133.791	-12.055	**-37.631**	-13.474	**-3.057**
Drugbank	**-32.461**	0	-22.638	-97.702	-31.323	-233.392	-16.743	-6.493
Repur	**-32.461**	-31.596	0	-100.765	-107.863	-72.401	-17.807	-9.283
PDX	-36.485	-38.241	**-17.59**	0	**79.375**	-403.209	-30.263	-13.403
GDSC	-39.56	-30.866	-23.169	**-61.647**	0	-59.743	-27.124	-21.273
CCLE	-41.195	-40.007	-18.691	-91.797	-185.611	0	-27.004	-12.479
SIDER	-33.511	-31.224	-22.36	-120.254	18.839	-211.632	0	-15.539
OFFSIDE	-33.511	-31.271	-22.518	-96.776	20.079	-62.449	**-10.611**	0

2.1.5 Gradient Boosting Tree

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	**-5.162**	**-9.322**	-139.546	-131.397	**-72.317**	**-8.893**	-3.007
Drugbank	-30.816	0	-16.208	-106.294	7.367	-185.569	-8.945	**6.475**
Repur	**-25.845**	-20.356	0	-93.892	117.343	-122.651	-18.571	4.649
PDX	-35.739	-18.649	-15.244	0	41.632	-256.766	-18.079	-0.73
GDSC	-33.657	-32.762	-18.239	**-81.142**	0	-180.29	-25.288	-21.964
CCLE	-38.31	-41.265	-14.488	-115.775	-116.173	0	-34.578	-11.505
SIDER	-37.331	-20.356	-15.146	-107.066	**199.707**	-176.308	0	-6.704
OFFSIDE	-32.065	-23.313	-17.219	-117.813	-9.724	-108.504	-8.945	0

2.1.6 SVM

Datasets	STITCH	Drugbank	Repur	PDX	GDSC	CCLE	SIDER	OFFSIDE
STITCH	0	-36.224	-24.242	-128.589	-76.179	-106.421	-25.595	-15.254
Drugbank	**-32.432**	0	-24.242	-150.93	**76.421**	-72.028	-25.595	-15.254
Repur	**-32.432**	-36.224	0	-111.049	-50.335	**-60.832**	-25.595	-15.254
PDX	-33.358	-35.489	**-21.322**	0	61.906	-179.947	**-17.27**	-7.468
GDSC	-35.234	**-32.65**	-31.83	**-72.177**	0	-89.026	-24.034	-16.66
CCLE	-41.688	-39.799	-25.567	-83.803	-80.256	0	-24.294	**-4.463**
SIDER	**-32.432**	-36.224	-24.242	-111.933	34.744	-67.119	0	-15.254
OFFSIDE	**-32.432**	-36.224	-24.242	-134.561	69.301	-146.474	-25.595	0

2.2 Tasks

2.2.1 KNN

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-47.474	-10.053
Drug Response	605.869	0	605.869
Drug Side Effects	17.101	21.316	0

2.2.2 Linear Regression

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-64.854	9.104
Drug Response	274.14	0	274.14
Drug Side Effects	52.861	-20.803	0

2.2.3 Decision Tree

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-37.739	19.596
Drug Response	799.749	0	799.749
Drug Side Effects	86.916	86.916	0

2.2.4 Random Forest

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-48.714	1.675
Drug Response	1350.736	0	1354.436
Drug Side Effects	50.29	13.964	0

2.2.5 Gradient Boosting Tree

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-53.852	-5.522
Drug Response	894.247	0	894.247
Drug Side Effects	29.251	1.621	0

2.2.6 SVM

Tasks	Drug Targets	Drug Response	Drug Side Effects
Drug Targets	0	-43.42	-12.139
Drug Response	334.703	0	334.703
Drug Side Effects	19.174	-20.699	0

3 GitHub

Link to GitHub: https://github.com/marendale/team-5-csc-4850