AUTOML – HOMEWORK 1

BARTOSZ JADCZAK, MACIEJ KORDYACZNY

DATASETY

Źródło danych openml

- **monks-problems-1, id: 333**
- **monks-problems-2, id: 334**
- **blood-transfusion-service-center, id: 1464**
- **Diabetes, id: 37**

ALGORYTMY I MODELE

Algorytmy:

- Random search
- Bayes optimization

Metody uczenia:

- Random forest
- Gradient boosting
- K-nearest neighbours

WYNIKI BAYES OPTIMALISATION

Bayes Optimalization						
	Params Set	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Mean
K-nearest neighbors (KNN)	1	0.76	0.7686	0.9643	0.7078	0.800175
	2	0.78	0.7934	0.9732	0.7468	0.82335
	3	0.76	0.7851	0.9554	0.7468	0.811825
	4	0.7467	0.7851	0.9286	0.7468	0.8018
	5	0.7467	0.7851	0.9464	0.7403	0.804625
	6	0.7467	0.7851	0.9554	0.7338	0.80525
Random forest	1	0.7597	0.7597	0.7532	0.7857	0.764575
	2	0.7597	0.7597	0.7338	0.7532	0.7516
	3	0.7792	0.7143	0.7662	0.7403	0.75
	4	0.7667	0.9504	1	0.7597	0.8692
	5	0.7867	0.9256	1	0.7662	0.869625
	6	0.7867	0.8264	1	0.7792	0.848075
Gradient Boosting	1	0.7533	1	1	0.7532	0.876625
	2	0.7933	1	1	0.7208	0.878525
	3	0.72	1	1	0.7403	0.865075
	4	0.7533	0.6281	0.6964	0.7857	0.715875
	5	0.7733	1	1	0.7662	0.884875
	6	0.7667	1	1	0.7468	0.878375

PORÓWNANIE WYNIKÓW

	Params Set	Dataset 1	Dataset 2	Dataset 3	Dataset 4
K-nearest neighbors (KNN)	1	-0.06335	-0.05475	0.14095	-0.11555
	2	-0.04335	-0.02995	0.14985	-0.07655
	3	-0.06335	-0.03825	0.13205	-0.07655
	4	-0.07665	-0.03825	0.10525	-0.07655
	5	-0.07665	-0.03825	0.12305	-0.08305
	6	-0.07665	-0.03825	0.13205	-0.08955
Random forest	1	-0.109925	-0.109925	-0.116425	-0.083925
	2	-0.109925	-0.109925	-0.135825	-0.116425
	3	-0.090425	-0.155325	-0.103425	-0.129325
	4	-0.102925	0.080775	0.130375	-0.109925
	5	-0.082925	0.055975	0.130375	-0.103425
	6	-0.082925	-0.043225	0.130375	-0.090425
Gradient Boosting	1	-0.131575	0.115125	0.115125	-0.131675
	2	-0.091575	0.115125	0.115125	-0.164075
	3	-0.164875	0.115125	0.115125	-0.144575
	4	-0.131575	-0.256775	-0.188475	-0.099175
	5	-0.111575	0.115125	0.115125	-0.118675
	6	-0.118175	0.115125	0.115125	-0.138075

WYNIKI RANDOM SEARCH

RANDOM FOREST

		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Mean
Random forest	1	0.7533	0.6281	0.6964	0.7403	0.704525
	2	0.7733	0.6281	0.7768	0.7468	0.73125
	3	0.7533	0.6281	0.6964	0.6429	0.680175
	4	0.64	0.9587	0.9732	0.7403	0.82805
	5	0.7533	0.6281	0.6964	0.7532	0.70775
	6	0.7533	0.6281	0.6964	0.7403	0.704525
	7	0.7533	0.6281	0.6964	0.6429	0.680175
	8	0.7	0.9817	1	0.7403	0.8555
	9	0.76	0.843	0.9554	0.7273	0.821425
	10	0.7133	0.9256	0.8661	0.7532	0.81455
	11	0.7267	0.9091	1	0.7208	0.83915
	12	0.7667	0.7521	0.9554	0.7338	0.802
	13	0.7333	0.9001	1	0.7403	0.843425
	14	0.7467	0.9008	1	0.7403	0.84695
	15	0.7533	0.6281	0.6964	0.6429	0.680175
	16	0.7533	0.6281	0.6964	0.6429	0.680175
	17	0.7533	0.6281	0.6964	0.6494	0.6818
	18	0.7533	0.6281	0.6964	0.6429	0.680175
	19	0.7533	0.6281	0.6964	0.6429	0.680175
	20	0.7333	0.8595	1	0.7532	0.8365

WYNIKI RANDOM SEARCH

GRADIENT BOOSTING

		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Mean
Gradient Boosting	21	0.7533	0.6281	0.6518	0.6429	0.669025
	22	0.7533	0.6281	0.4464	0.6429	0.617675
	23	0.7533	0.6281	0.6964	0.7662	0.711
	24	0.7533	0.6281	0.6518	0.6429	0.669025
	25	0.7533	0.6281	0.5714	0.6429	0.648925
	26	0.7533	0.6281	0.6518	0.7078	0.68525
	27	0.7533	0.6281	0.4464	0.6429	0.617675
	28	0.7533	0.6281	0.4464	0.6429	0.617675
	29	0.7533	0.6281	0.4464	0.6429	0.617675
	30	0.7533	0.6281	0.4464	0.6429	0.617675
	31	0.7533	0.6281	0.4464	0.6429	0.617675
	32	0.7533	0.6281	0.4464	0.6429	0.617675
	33	0.7533	0.6281	0.6518	0.6429	0.669025
	34	0.7533	0.6281	0.4464	0.6429	0.617675
	35	0.7533	0.6281	0.6964	0.6429	0.680175
	36	0.7533	0.6281	0.6518	0.6429	0.669025
	37	0.7533	0.6281	0.4464	0.6429	0.617675
	38	0.7533	0.6281	0.6518	0.7208	0.6885
	39	0.7533	0.6281	0.6964	0.6429	0.680175
	40	0.7533	0.6281	0.4464	0.6429	0.617675

WYNIKI RANDOM SEARCH

KNN

		Dataset 1	Dataset 2	Dataset 3	Dataset 4	Mean
K-nearest neighbors (KNN)	41	0.7133	0.7025	0.9375	0.6688	0.755525
	42	0.7733	0.6942	0.9643	0.6883	0.780025
	43	0.7733	0.6281	0.8214	0.6883	0.727775
	44	0.78	0.6446	0.9643	0.7403	0.7823
	45	0.7	0.719	0.9554	0.6753	0.762425
	46	0.7067	0.595	0.9464	0.6688	0.729225
	47	0.7733	0.6777	0.8482	0.6883	0.746875
	48	0.78	0.6446	0.9643	0.7403	0.7823
	49	0.7	0.7438	0.9732	0.6883	0.776325
	50	0.76	0.7603	0.9554	0.7078	0.795875
	51	0.7533	0.6942	0.9554	0.6883	0.7728
	52	0.7	0.719	0.9643	0.6883	0.7679
	53	0.6867	0.5702	0.9643	0.6883	0.727375
	54	0.76	0.7686	0.9732	0.7078	0.8024
	55	0.76	0.6281	0.8214	0.6494	0.714725
	56	0.7	0.7025	0.8393	0.6494	0.7228
	57	0.7133	0.7355	0.875	0.6623	0.746525
	58	0.7333	0.7686	0.8839	0.6948	0.77015
	59	0.6733	0.7769	0.8839	0.6948	0.757225
	60	0.6867	0.5702	0.9643	0.6883	0.727375

PORÓWNANIE WYNIKÓW

		Dataset 1	Dataset 2	Dataset 3	Dataset 4
K-nearest neighbors (KNN)	1	-0.1022	-0.2274	-0.1591	-0.1152
	2	-0.0822	-0.2274	-0.0787	-0.1087
	3	-0.1022	-0.2274	-0.1591	-0.2126
	4	-0.2155	0.1032	0.1177	-0.1152
	5	-0.1022	-0.2274	-0.1591	-0.1023
	6	-0.1022	-0.2274	-0.1591	-0.1152
	7	-0.1022	-0.2274	-0.1591	-0.2126
	8	-0.1555	0.1262	0.1445	-0.1152
	9	-0.0955	-0.0125	0.0999	-0.1282
	10	-0.1422	0.0701	0.0106	-0.1023
	11	-0.1288	0.0536	0.1445	-0.1347
	12	-0.0888	-0.1034	0.0999	-0.1217
	13	-0.1222	0.0446	0.1445	-0.1152
	14	-0.1088	0.0453	0.1445	-0.1152
	15	-0.1022	-0.2274	-0.1591	-0.2126
	16	-0.1022	-0.2274	-0.1591	-0.2126
	17	-0.1022	-0.2274	-0.1591	-0.2061
	18	-0.1022	-0.2274	-0.1591	-0.2126
	19	-0.1022	-0.2274	-0.1591	-0.2126
	20	-0.1222	0.004	0.1445	-0.1023

PORÓWNANIE WYNIKÓW

		Dataset 1	Dataset 2	Dataset 3	Dataset 4
Random forest	21	0.0423	-0.0829	-0.0592	-0.0681
	22	0.0423	-0.0829	-0.2646	-0.0681
	23	0.0423	-0.0829	-0.0146	0.0552
	24	0.0423	-0.0829	-0.0592	-0.0681
	25	0.0423	-0.0829	-0.1396	-0.0681
	26	0.0423	-0.0829	-0.0592	-0.0032
	27	0.0423	-0.0829	-0.2646	-0.0681
	28	0.0423	-0.0829	-0.2646	-0.0681
	29	0.0423	-0.0829	-0.2646	-0.0681
	30	0.0423	-0.0829	-0.2646	-0.0681
	31	0.0423	-0.0829	-0.2646	-0.0681
	32	0.0423	-0.0829	-0.2646	-0.0681
	33	0.0423	-0.0829	-0.0592	-0.0681
	34	0.0423	-0.0829	-0.2646	-0.0681
	35	0.0423	-0.0829	-0.0146	-0.0681
	36	0.0423	-0.0829	-0.0592	-0.0681
	37	0.0423	-0.0829	-0.2646	-0.0681
	38	0.0423	-0.0829	-0.0592	0.0098
	39	0.0423	-0.0829	-0.0146	-0.0681
	40	0.0423	-0.0829	-0.2646	-0.0681

PORÓWNANIE WYNIKÓW

		Dataset 1	Dataset 2	Dataset 3	Dataset 4
Gradient boosting	41	-0.0891	-0.0999	0.1351	-0.1336
	42	-0.0291	-0.1082	0.1619	-0.1141
	43	-0.0291	-0.1743	0.019	-0.1141
	44	-0.0224	-0.1578	0.1619	-0.0621
	45	-0.1024	-0.0834	0.153	-0.1271
	46	-0.0957	-0.2074	0.144	-0.1336
	47	-0.0291	-0.1247	0.0458	-0.1141
	48	-0.0224	-0.1578	0.1619	-0.0621
	49	-0.1024	-0.0586	0.1708	-0.1141
	50	-0.0424	-0.0421	0.153	-0.0946
	51	-0.0491	-0.1082	0.153	-0.1141
	52	-0.1024	-0.0834	0.1619	-0.1141
	53	-0.1157	-0.2322	0.1619	-0.1141
	54	-0.0424	-0.0338	0.1708	-0.0946
	55	-0.0424	-0.1743	0.019	-0.153
	56	-0.1024	-0.0999	0.0369	-0.153
	57	-0.0891	-0.0669	0.0726	-0.1401
	58	-0.0691	-0.0338	0.0815	-0.1076
	59	-0.1291	-0.0255	0.0815	-0.1076
	60	-0.1157	-0.2322	0.1619	-0.1141