Table 1 : Pipeline structure and accuracy of the best pipeline, number of generated pipelines, pipeline evaluation time and wasted evaluation time generated by AutoWeka4MCPS using different datasets and seed numbers

dataset	criteria	seed=0	seed=1	seed=2	seed=3	seed=4
abalone	structure of the best pipeline	trees.LMT	bayes.BayesNet	functions. MultilayerPerceptron	attributeSelection.Ranker → functions.SimpleLogistic	attributeSelection.BestFirst → trees.LMT
	accuracy of the best pipeline (%)	25.62	26.02	**26.18**	25.62	25.94
	invalid/ total evaluated pipelines	16/42	90/169	69/157	34/63	53/133
	average evaluation time of invalid/ total pipelines (ms)	225,479/114,655	22,301/19,079	65,404/42,311	106,334/60,827	437/26,360
	wasted evaluation time (%)	73.18	61.88	67.51	92.87	0.66
car	structure of the best pipeline	attributeSelection.BestFirst → functions.SMO	trees.J48	attributeSelection.BestFirst → functions.SMO	trees.LMT	functions.SMO
	accuracy of the best pipeline (%)	**100.00**	96.14	**100.00**	97.68	97.68
	invalid/ total evaluated pipelines	205/357	108/178	197/510	139/295	85/149
	average evaluation time of invalid/ total pipelines (ms)	18,625/11,480	32,394/20,176	22,962/9,895	37,642/18,525	51,355/29,701
	wasted evaluation time (%)	92.90	96.87	89.47	95.42	97.98
convex	structure of the best pipeline	trees.J48	-	attribute.Center → attribute.PrincipalComponents → instance.PeriodicSampling → meta.ClassificationViaRegression	crashed	crashed
	accuracy of the best pipeline (%)	**62.44**	-	50.29	crashed	crashed
	invalid/ total evaluated pipelines	18/38	2/2	17/28	crashed	crashed
	average evaluation time of invalid/ total pipelines (ms)	4,239/93,959	1,737,585/1,158,390	77,921/126,083	crashed	crashed
	wasted evaluation time (%)	2.08	100.00	36.23	crashed	crashed
gcredit	structure of the best pipeline	bayes.NaiveBayes	functions. MultilayerPerceptron	functions.SMO	bayes.NaiveBayes	attributeSelection.Ranker → functions.Logistic
	accuracy of the best pipeline (%)	69.33	50.05	70.33	69.33	**73.67**
	invalid/ total evaluated pipelines	112/307	229/593	208/374	12/66	30/84
	average evaluation time of invalid/ total pipelines (ms)	25,187/16,497	16,724/6,928	18,912/10,981	305,632/55,248	121,160/43,524
	wasted evaluation time (%)	55.52	93.06	95.53	99.08	98.25
wineqw	structure of the best pipeline	lazy.Kstar	functions.SMO	crashed	lazy.Kstar	trees.J48
	accuracy of the best pipeline (%)	**64.81**	59.7	crashed	64.33	59.09
	invalid/ total evaluated pipelines	203/416	121/260	crashed	201/503	36/90
	average evaluation time of invalid/ total pipelines (ms)	24,042/14,229	34,573/20,161	crashed	12,032/7,045	45,534/27,489
	wasted evaluation time (%)	82.26	79.50	crashed	68.11	65.53

Table 2 : Pipeline structure and accuracy of the best pipeline, number of generated pipelines, pipeline evaluation time and wasted evaluation time generated by Auto-sklearn using different datasets and seed numbers

dataset	criteria	seed=0	seed=1	seed=2	seed=3	seed=4
abalone	structure of the best pipeline	crashed	crashed	crashed	crashed	crashed
	accuracy of the best pipeline (%)					
	invalid/total evaluated pipelines					
	average evaluation time of invalid/ total pipelines (ms)					
	wasted evaluation time (%)					
car	structure of the best pipeline	crashed	crashed	crashed	crashed	crashed
	accuracy of the best pipeline (%)					
	invalid/total evaluated pipelines					
	average evaluation time of invalid/ total pipelines (ms)					
	wasted evaluation time (%)					
convex	structure of the best pipeline	OneHotEncoding→Median →FeatureAgglomeration →QuantileTransformer →GradientBoosting	OneHotEncoding →Median →FeatureAgglomeration →QuantileTransformer →GradientBoosting	OneHotEncoding →Median →FeatureAgglomeration →QuantileTransformer →GradientBoosting	OneHotEncoding→Median →FeatureAgglomeration →QuantileTransformer →GradientBoosting	OneHotEncoding→Median →FeatureAgglomeration →QuantileTransformer →GradientBoosting
	accuracy of the best pipeline (%)	83.36	81.84	**83.42**	82.73	82.65
	invalid/total evaluated pipelines	2/18	2/15	2/12	2/11	2/14
	average evaluation time of invalid/ total pipelines (ms)	28,040/19,770	26,884/23,791	29,203/29,695	27,902/32,395	27,999/25,451
	wasted evaluation time (%)	15.76	15.07	16.39	15.66	15.72
gcredit	structure of the best pipeline	crashed	crashed	crashed	crashed	crashed
	accuracy of the best pipeline (%)					
	invalid/total evaluated pipelines					
	average evaluation time of invalid/ total pipelines (ms)					
	wasted evaluation time (%)					
wineqw	structure of the best pipeline	BalancingWeighting →MostFreqPoly →RobustScaler →Adaboost	BalancingWeighting →MostFreqPoly →RobustScaler →Adaboost	BalancingWeighting →MostFreqPoly →RobustScaler →Adaboost	BalancingWeighting →MostFreqPoly →RobustScaler →Adaboost	BalancingWeighting →MostFreqPoly →RobustScaler →Adaboost
	accuracy of the best pipeline (%)	**65.83**	**65.83**	**65.83**	**65.83**	**65.83**
	invalid/ total evaluated pipelines	0/43	0/24	0/43	0/33	0/33
	average evaluation time of invalid/ total pipelines (ms)	0/8,280	0/14,915	0/8,277	0/10,815	0/10,813
	wasted evaluation time (%)	0.00	0.00	0.00	0.00	0.00