

Binary Classification Datasets

OpenML ID	Name	Imputation Target Column	Data Points	Numeric Features	Categorical Features
737	space_ga	INCOME	3107.0	6.0	1.0
871	pollen	RIDGE	3848.0	5.0	1.0
40983	wilt	Mean_G	4839.0	5.0	1.0
728	analcata_data_supreme	Lower_court_disagreement	4052.0	7.0	1.0
1489	phoneme	V1	5404.0	5.0	1.0
803	delta_aileron	RollRate	7129.0	5.0	1.0
923	visualizing_soil	isns	8641.0	3.0	2.0
725	bank8FM	a2pop	8192.0	8.0	1.0
42192	compas-two-years	age	5278.0	7.0	7.0
1558	bank-marketing	V7	4521.0	7.0	10.0
310	mammography	attr2	11183.0	6.0	1.0
1046	mozilla4	end	15545.0	5.0	1.0
847	wind	year	6574.0	14.0	1.0
40701	churn	total_day_minutes	5000.0	16.0	5.0
41146	sylvine	V4	5124.0	20.0	1.0
1496	ringnorm	V4	7400.0	20.0	1.0
1507	twonorm	V5	7400.0	20.0	1.0
823	houses	total_rooms	20640.0	8.0	1.0
42493	airlines	Lenght	26969.0	2.0	6.0
1471	eeg-eye-state	V9	14980.0	14.0	1.0
1120	MagicTelescope	fAlpha:	19020.0	11.0	1.0
4135	Amazon_employee_access	MGR_ID	32769.0	0.0	10.0
137	BNG(tic-tac-toe)	top-middle-square	39366.0	0.0	10.0
251	BNG(breast-w)	Cell_Shape_Uniformity	39366.0	9.0	1.0
1220	Click_prediction_small	keyword_id	39948.0	9.0	1.0
151	electricity	nswprice	45312.0	7.0	2.0
901	fried	X2	40768.0	10.0	1.0
881	mv	x3	40768.0	7.0	4.0
40922	Run_or_walk_information	gyro_y	88588.0	6.0	1.0
42477	default-of-credit-card-clients	x1	30000.0	23.0	1.0
23517	numera128.6	attribute_10	96320.0	21.0	1.0

Multiclass Classification Datasets

OpenML ID	Name	Imputation Target Column	Data Points	Numeric Features	Categorical Features
183	abalone	Length	4177.0	7.0	2.0
40498	wine-quality-white	V11	4898.0	11.0	1.0
30	page-blocks	eccen	5473.0	10.0	1.0
40677	led24	attribute#3	3200.0	0.0	25.0
1459	artificial-characters	V7	10218.0	7.0	1.0
40497	thyroid-ann	V13	3772.0	21.0	1.0
4552	BachChoralHarmony	V2	5665.0	2.0	15.0
26	nursery	parents	12960.0	0.0	9.0
375	JapaneseVowels	coefficient3	9961.0	14.0	1.0
32	pendigits	input4	10992.0	16.0	1.0
1481	kr-vs-k	V2	28056.0	3.0	4.0
184	kropt	black_king_col	28056.0	0.0	7.0
41027	jungle_chess_2pcs_raw_endgame_complete	black_piece0_strength	44819.0	6.0	1.0
6	letter	x-box	20000.0	16.0	1.0
41671	microaggregation2	a9	20000.0	20.0	1.0
40685	shuttle	A2	58000.0	9.0	1.0

Regression Datasets

OpenML ID	Name	Imputation Target Column	Data Points	Numeric Features	Categorical Features
198	delta_elevators	curRoll	9517.0	7.0	0.0
23515	sulfur	a3	10081.0	7.0	0.0
189	kin8nm	theta8	8192.0	9.0	0.0
287	wine_quality	sulphates	6497.0	12.0	0.0
42636	Long	x18	4477.0	20.0	0.0
42688	Brazilian_houses	area	10692.0	9.0	4.0
42183	dataset_sales	day	10738.0	15.0	0.0
1199	BNG(echoMonths)	wall_index	17496.0	7.0	3.0
197	cpu_act	pgin	8192.0	22.0	0.0
218	house_8L	P3	22784.0	9.0	0.0
42712	Bike_Sharing_Demand	humidity	17379.0	9.0	4.0
1193	BNG(lowbwt)	UI	31104.0	3.0	7.0
216	elevators	climbRate	16599.0	19.0	0.0
215	2dplanes	x6	40768.0	11.0	0.0
23395	COMET_MC_SAMPLE	wire_id	89640.0	6.0	0.0
42225	diamonds	carat	53940.0	7.0	3.0
1200	BNG(stock)	company1	59049.0	10.0	0.0
1213	BNG(mv)	x5	78732.0	8.0	3.0