Supplementary Material
for
What Works Best When? A Systematic Evaluation of Heuristics for Max-Cut and QUBO

Iain Dunning, Swati Gupta, John Silberholz

In this supplementary material, we give the CART models for each heuristic as described in Section 5.1, in the main paper. Table 1 summarizes the R^2 value of each fitted CART model and gives the list of relevant plots for all the 37 heuristics.

Heuristic	R^2	Figures	Heuristic	R^2	Figures	Heuristic	R^2	Figures
ALK98	0.52	Figure 1	BEA98SA	0.38	Figure 2	BEA98TS	0.74	Figure 3
BUR02	0.58	Figure 4	DUA05	0.54	Figure 5	FES02G	0.78	Figure 6
FES02GP	0.76	Figure 7	FES02GV	0.58	Figure 8	FES02GVP	0.68	Figure 9
FES02V	0.30	Figure 10	FES02VP	0.22	Figure 11	GLO10	0.49	Figure 12
GLO98	0.54	Figure 13	HAS00GA	0.52	Figure 14	HAS00TS	0.49	Figure 15
KAT00	0.38	Figure 16	KAT01	0.37	Figure 17	LAG09CE	0.19	Figure 18
LAG09HCE	0.48	Figure 19	LOD99	0.48	Figure 20	LU10	0.67	Figure 21
MER02GR	0.76	Figure 22	MER02GRK	0.46	Figure 23	MER02LS1	0.51	Figure 24
MER02LSK	0.54	Figure 25	MER04	0.31	Figure 26	MER99CR	0.46	Figure 27
MER99LS	0.45	Figure 28	MER99MU	0.21	Figure 29	PAL04MT	0.64	Figure 30
PAL04T1	0.72	Figure 31	PAL04T2	0.25	Figure 32	PAL04T3	0.29	Figure 33
PAL04T4	0.71	Figure 34	PAL04T5	0.65	Figure 35	PAL06	0.57	Figure 36
PAR08	0.61	Figure 37						

Table 1: The R^2 and figure number for each heuristic's CART model predicting instance-specific performance, as described in Section 5.1.

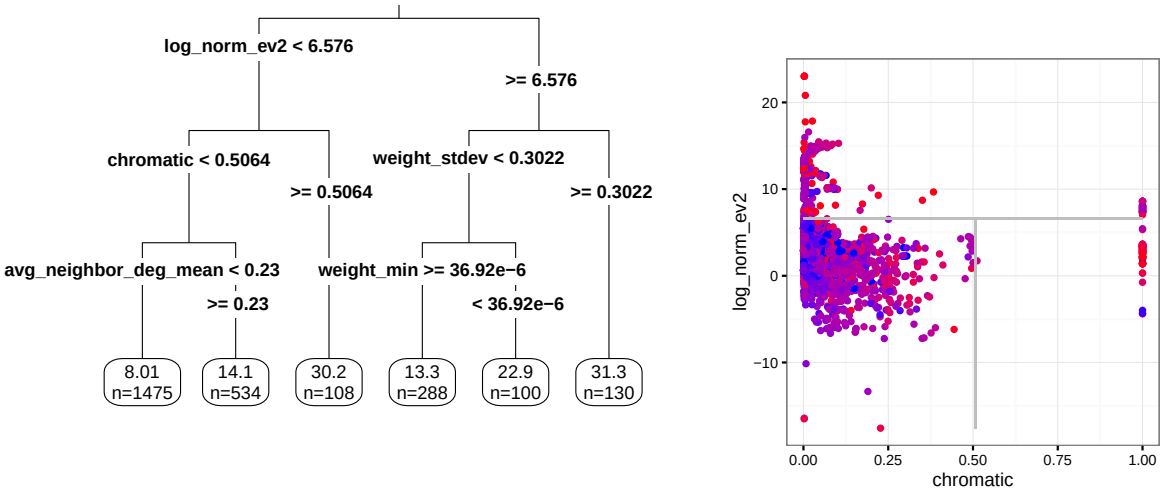

Figure 1: A CART model identifying instances on which ALK98 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

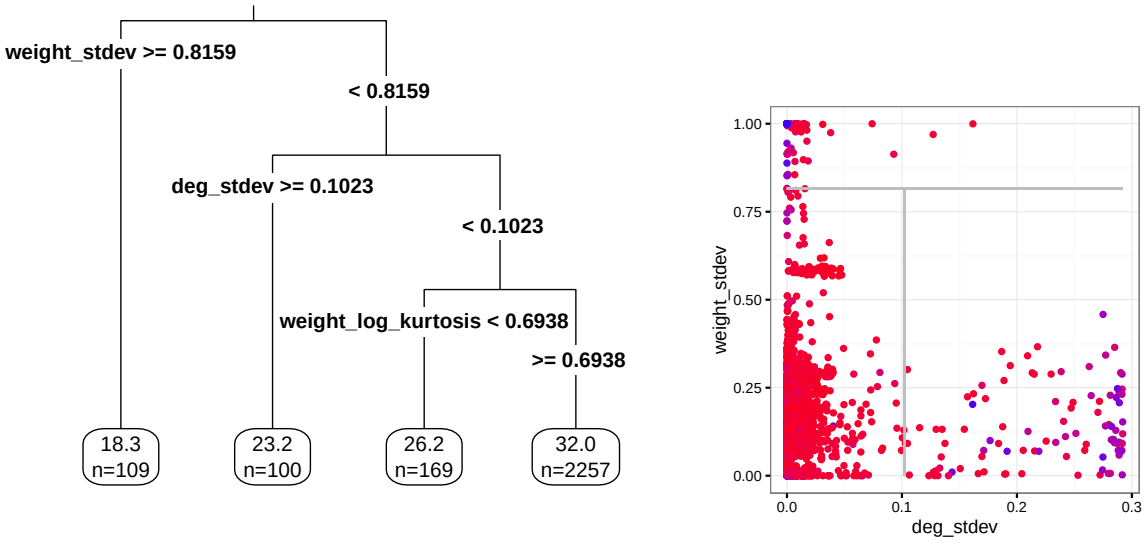

Figure 2: A CART model identifying instances on which BEA98SA performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

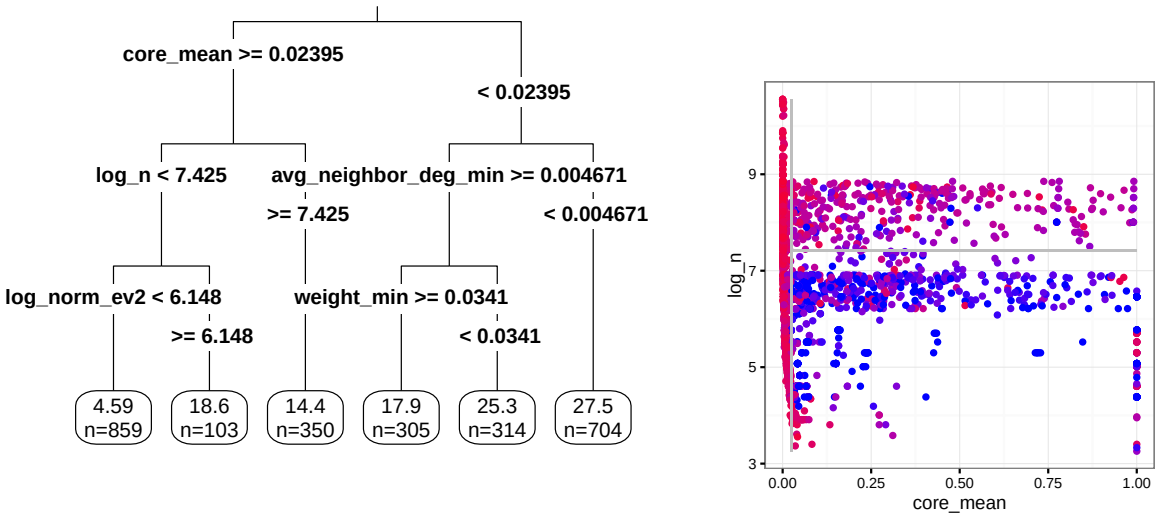

Figure 3: A CART model identifying instances on which BEA98TS performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

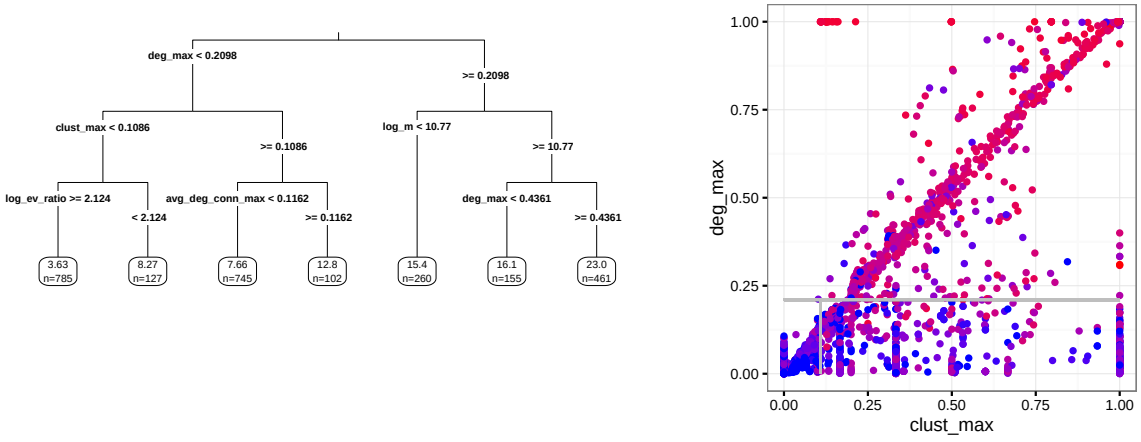

Figure 4: A CART model identifying instances on which BUR02 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

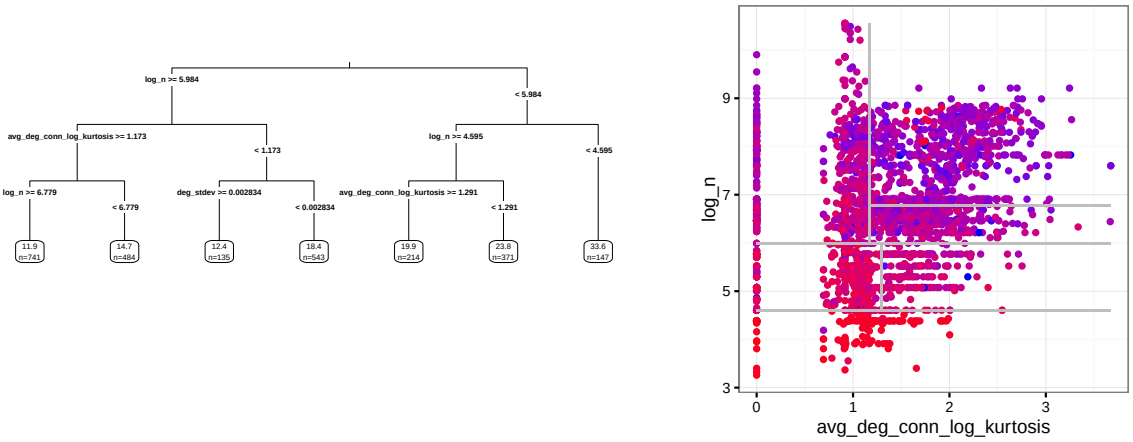

Figure 5: A CART model identifying instances on which DUA05 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

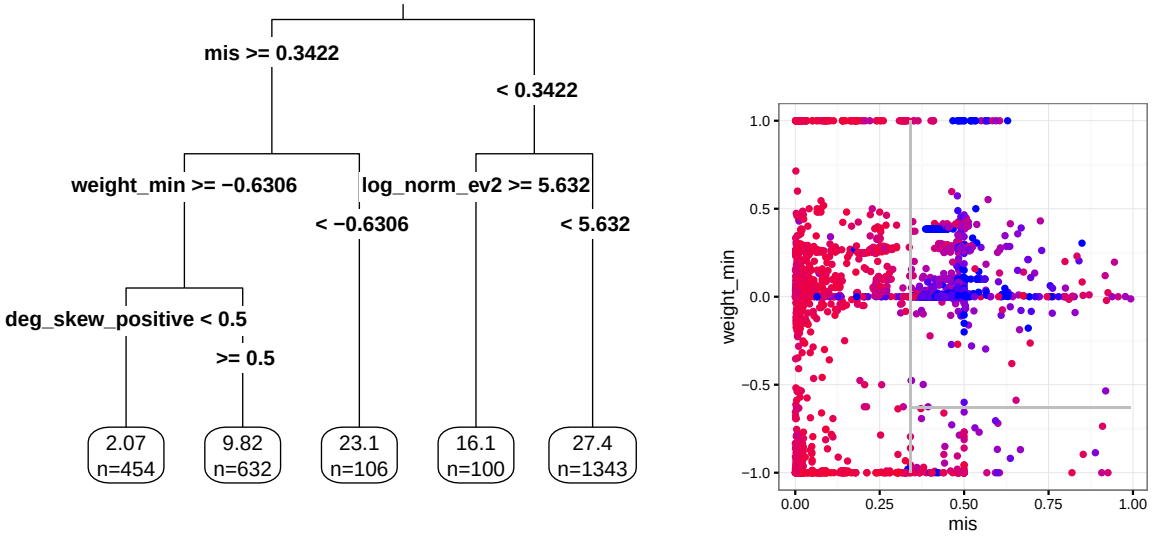

Figure 6: A CART model identifying instances on which FES02G performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

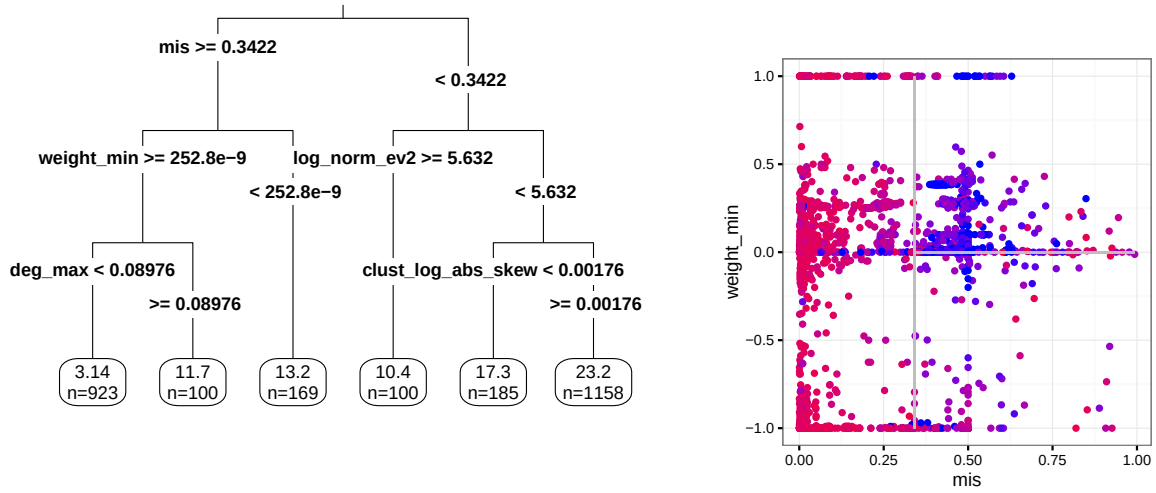

Figure 7: A CART model identifying instances on which FES02GP performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

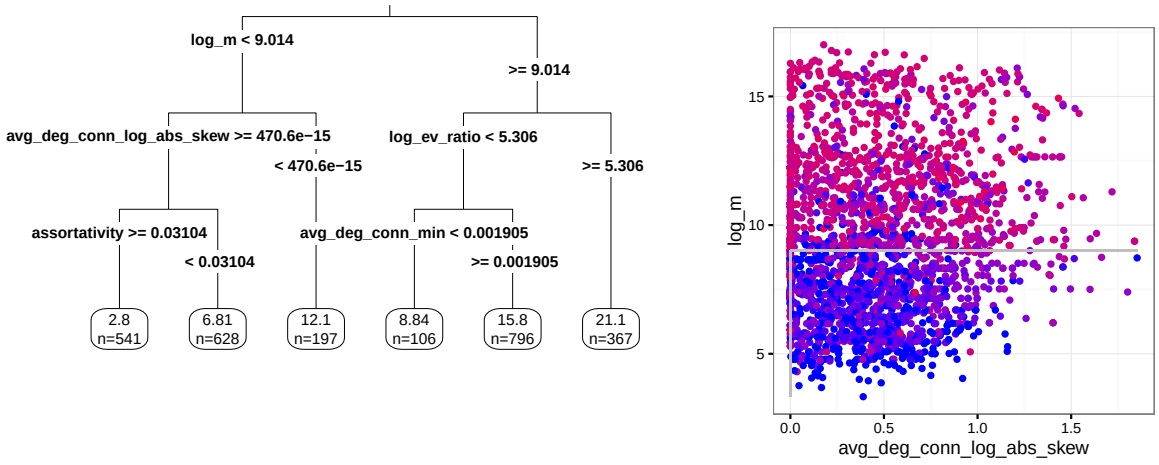

Figure 8: A CART model identifying instances on which FES02GV performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

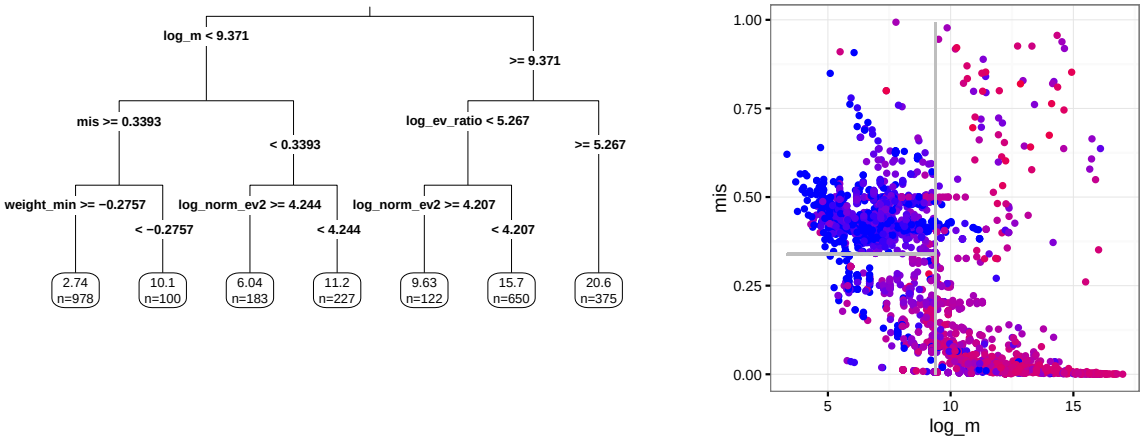

Figure 9: A CART model identifying instances on which FES02GVP performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

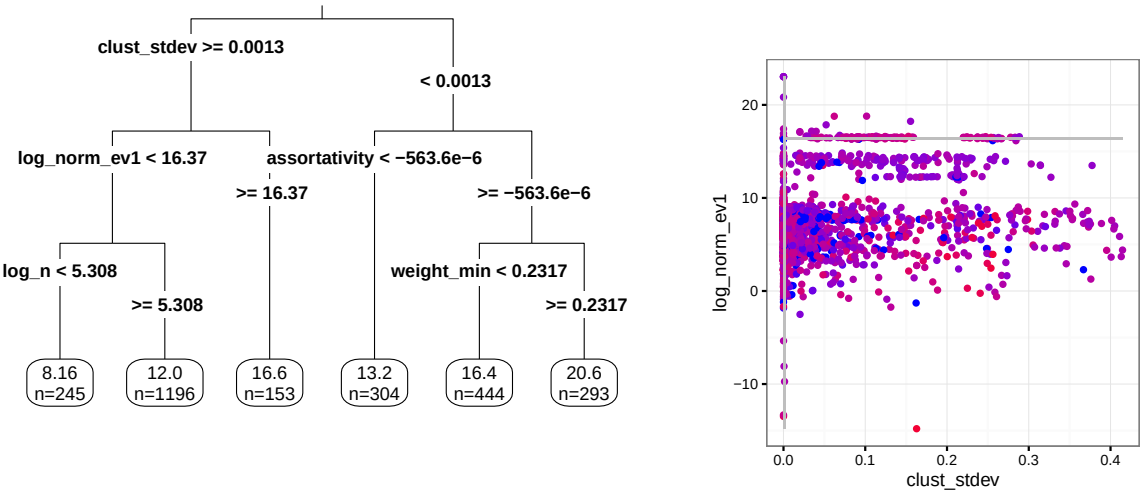

Figure 10: A CART model identifying instances on which FES02V performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

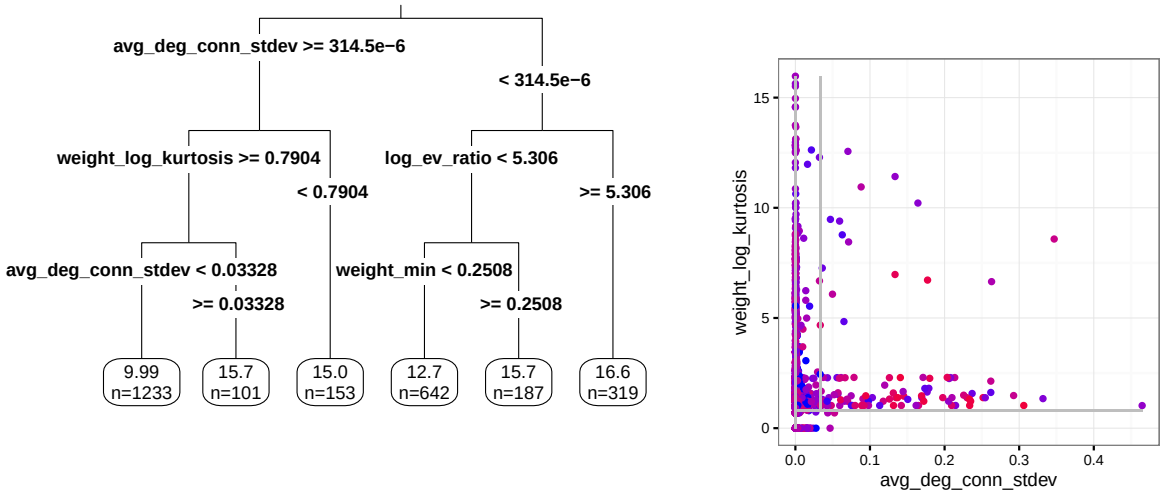

Figure 11: A CART model identifying instances on which FES02VP performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

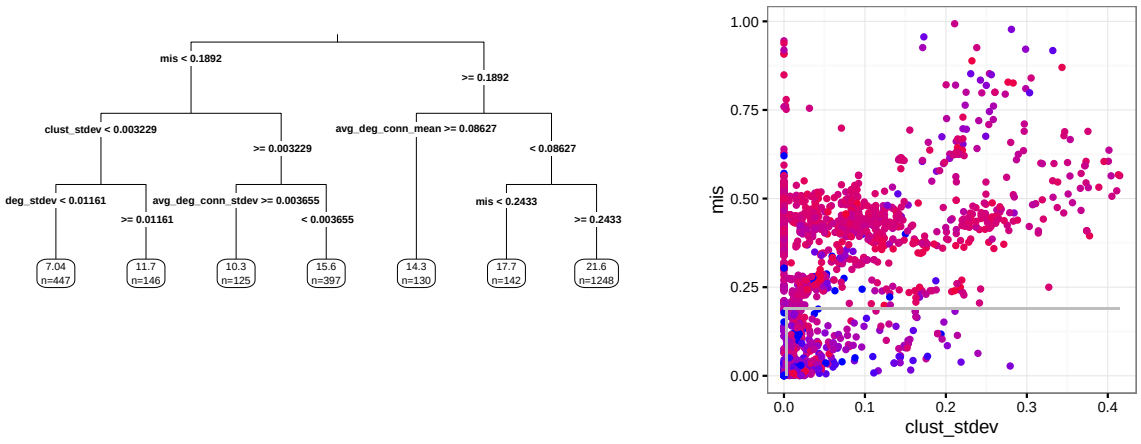

Figure 12: A CART model identifying instances on which GLO10 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

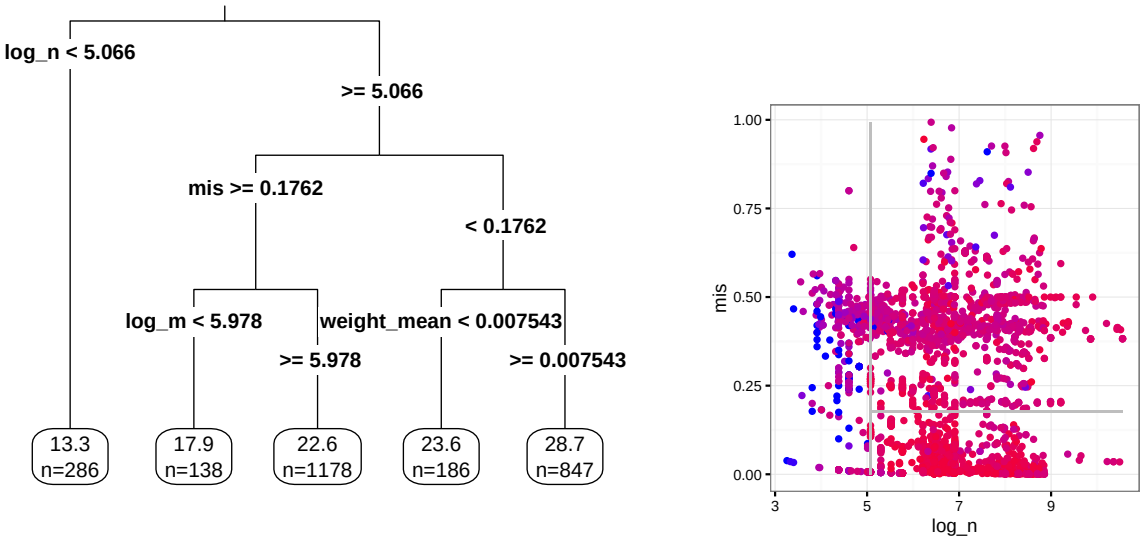

Figure 13: A CART model identifying instances on which GLO98 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

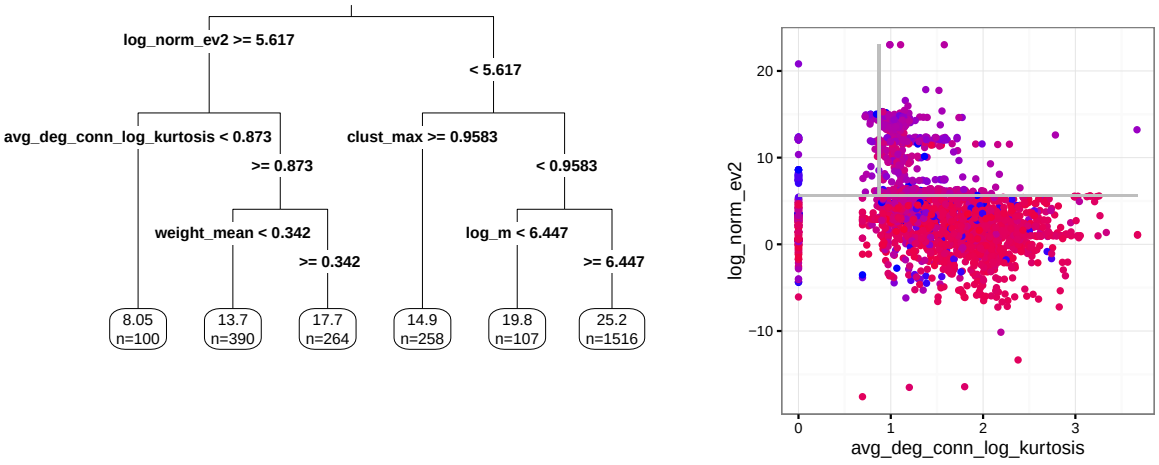

Figure 14: A CART model identifying instances on which HAS00GA performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

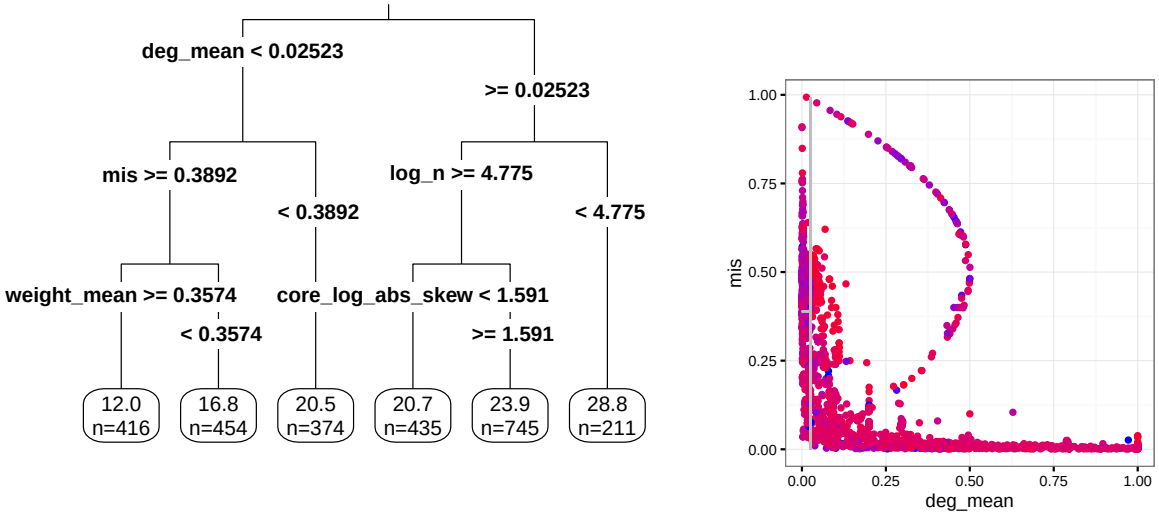

Figure 15: A CART model identifying instances on which HAS00TS performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

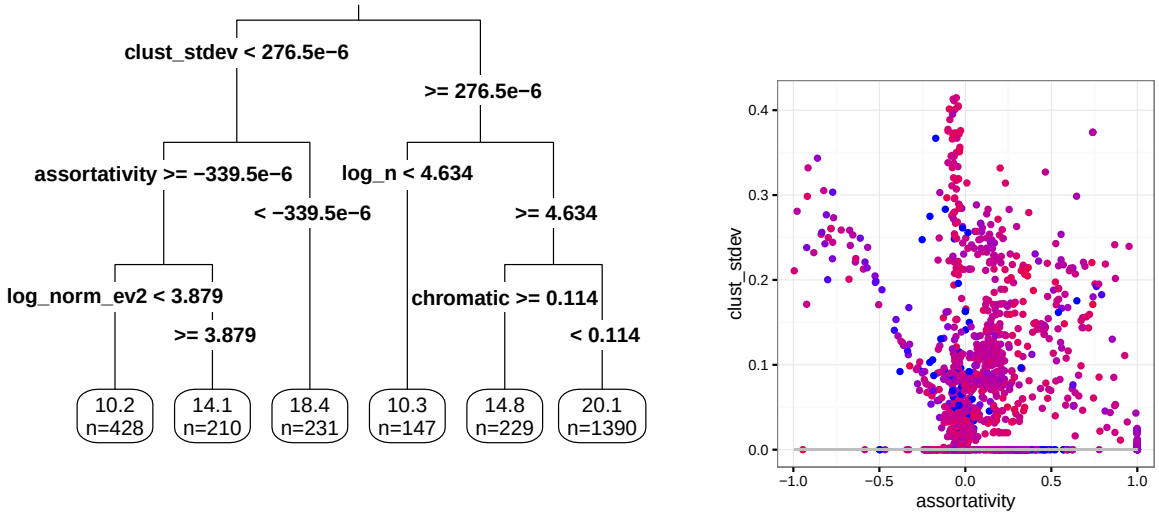

Figure 16: A CART model identifying instances on which KAT00 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

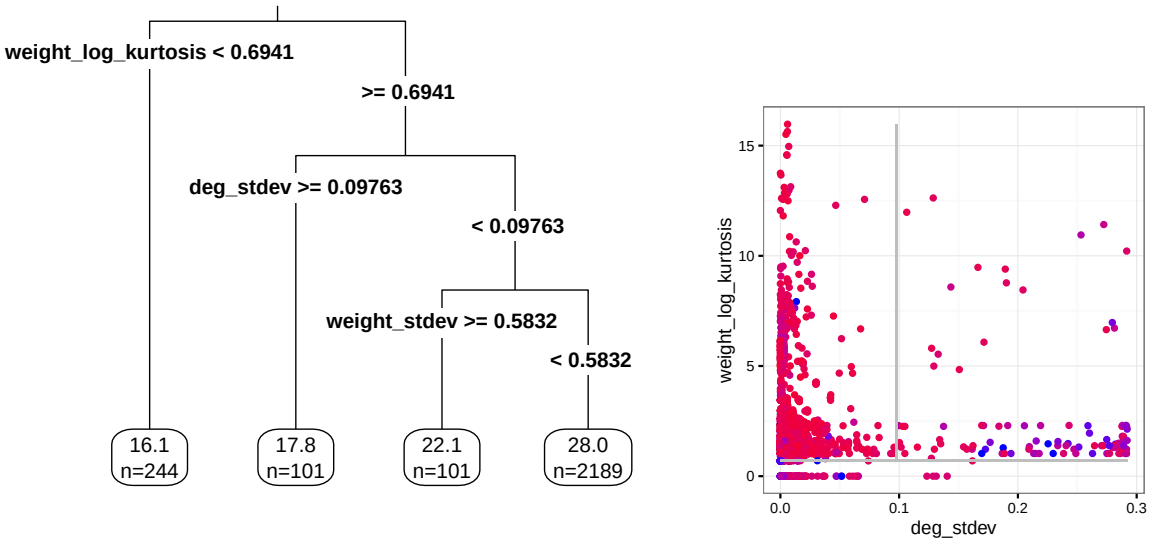

Figure 17: A CART model identifying instances on which KAT01 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

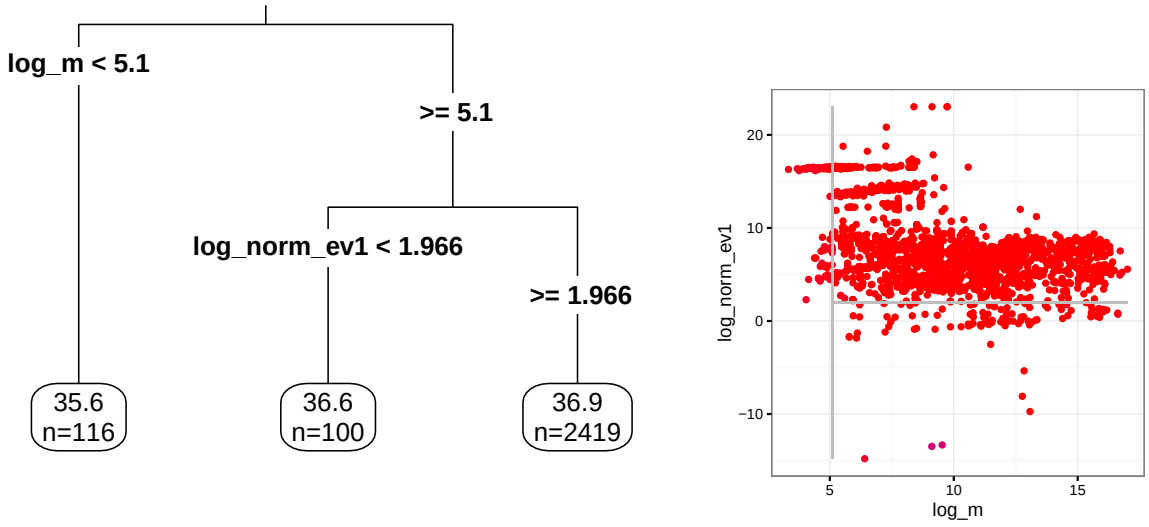

Figure 18: A CART model identifying instances on which LAG09CE performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

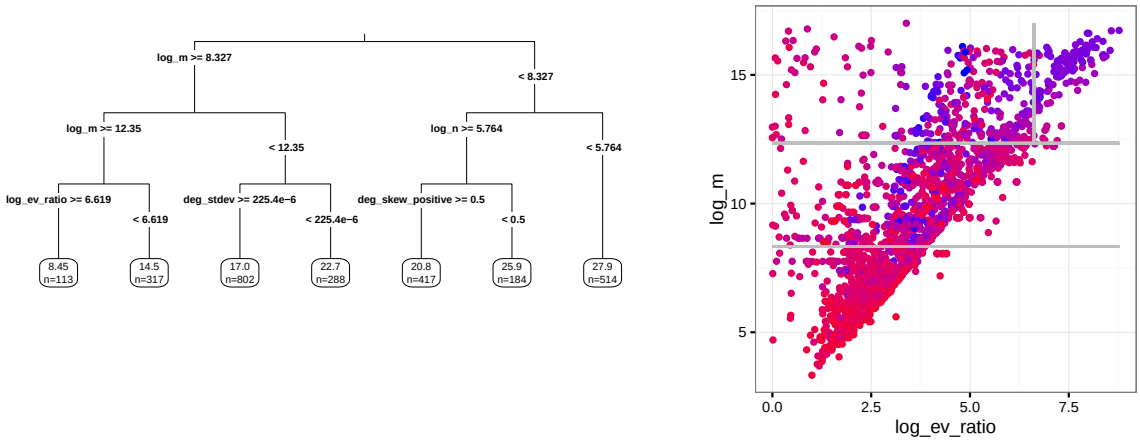

Figure 19: A CART model identifying instances on which LAG09HCE performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

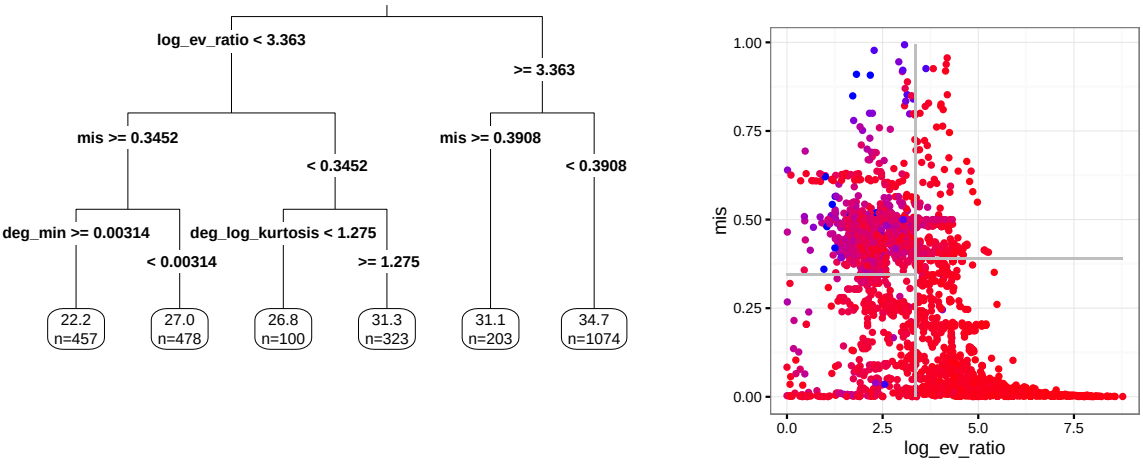

Figure 20: A CART model identifying instances on which LOD99 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

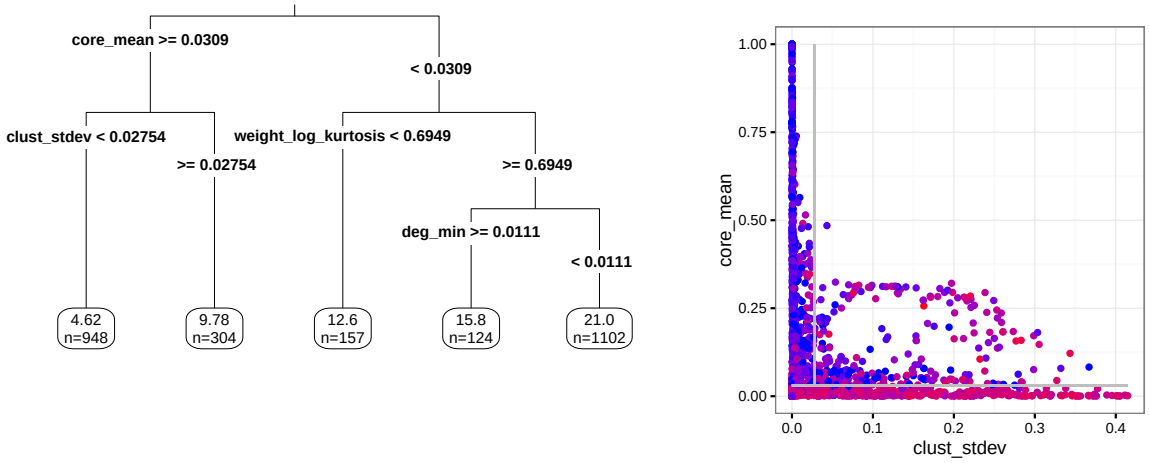

Figure 21: A CART model identifying instances on which LU10 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

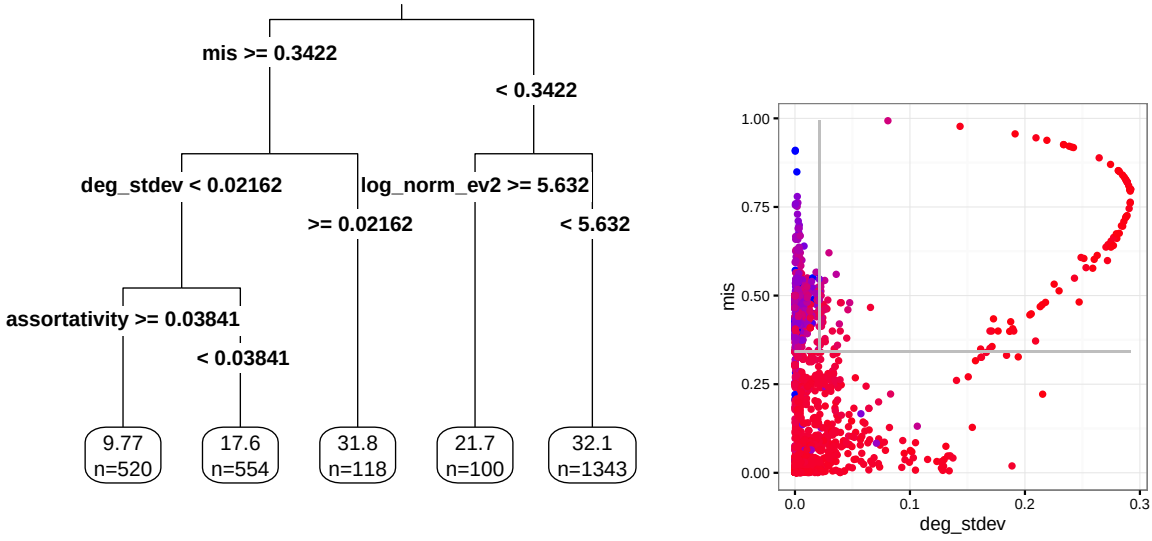

Figure 22: A CART model identifying instances on which MER02GR performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

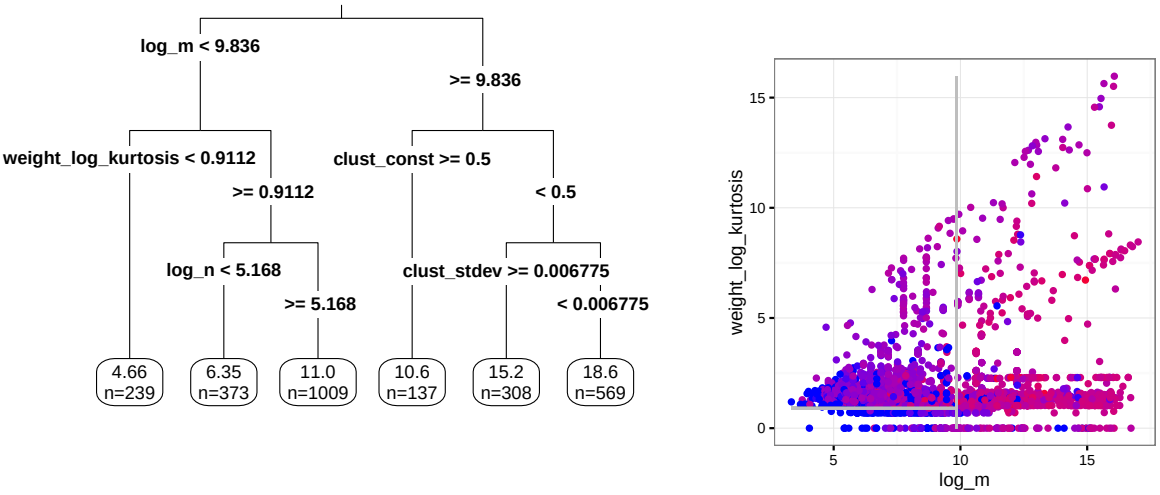

Figure 23: A CART model identifying instances on which MER02GRK performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

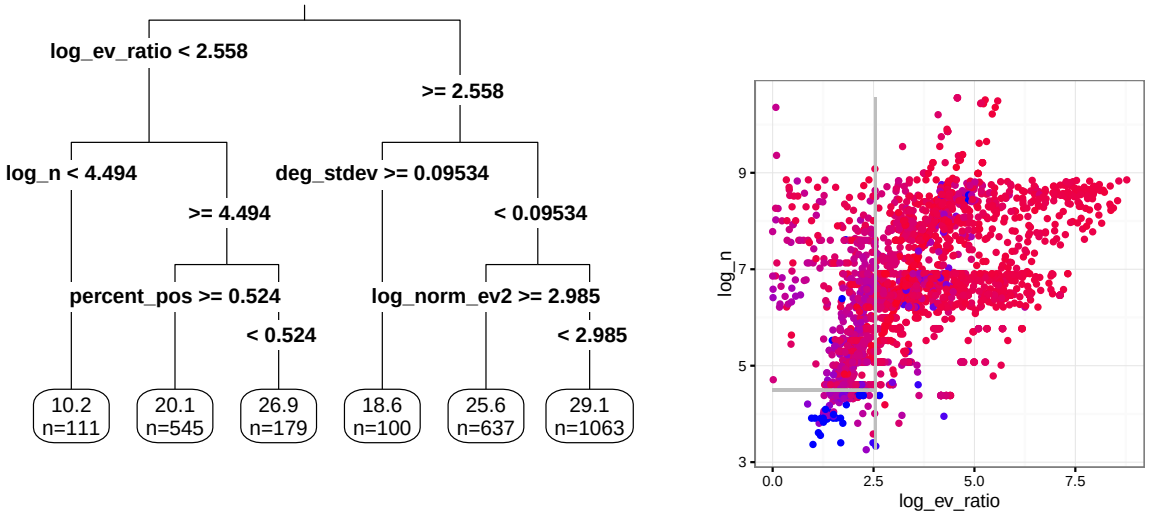

Figure 24: A CART model identifying instances on which MER02LS1 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

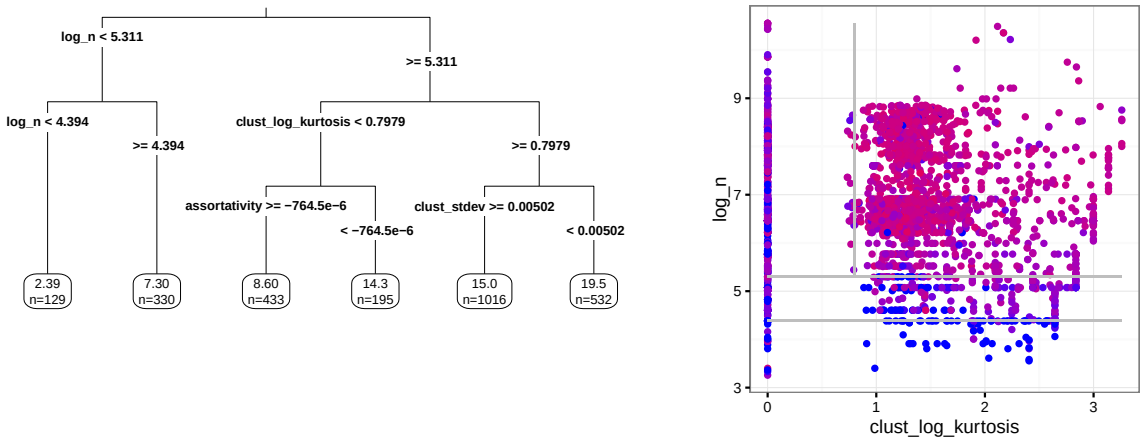

Figure 25: A CART model identifying instances on which MER02LSK performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

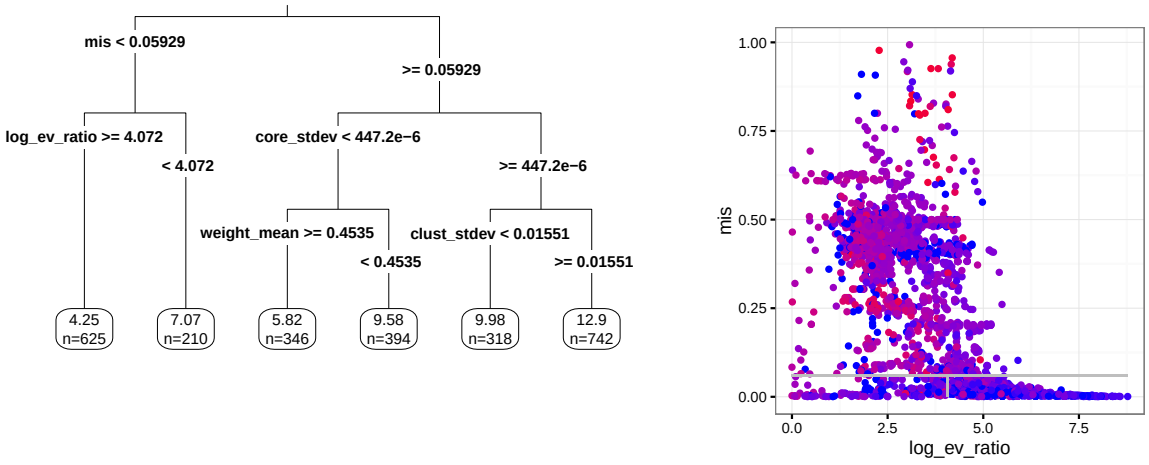

Figure 26: A CART model identifying instances on which MER04 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

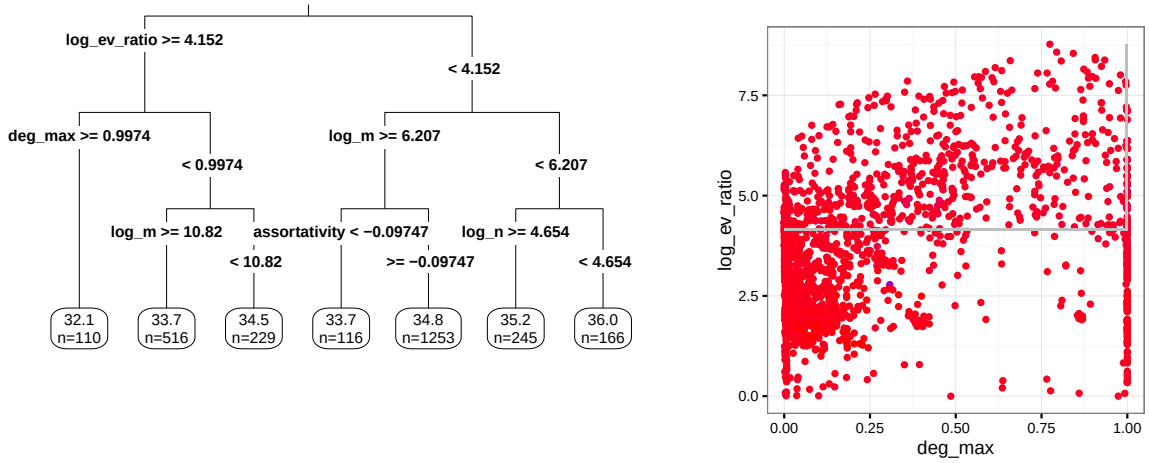

Figure 27: A CART model identifying instances on which MER99CR performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

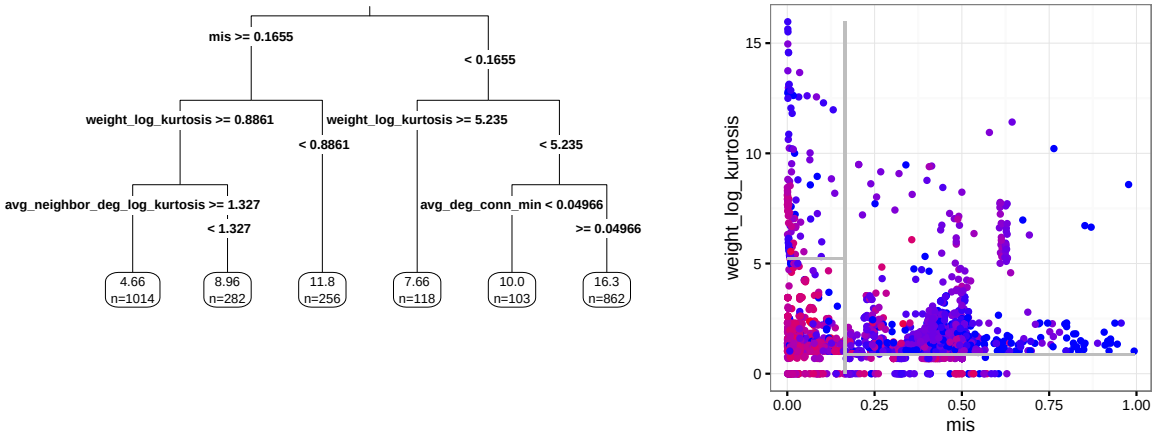

Figure 28: A CART model identifying instances on which MER99LS performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

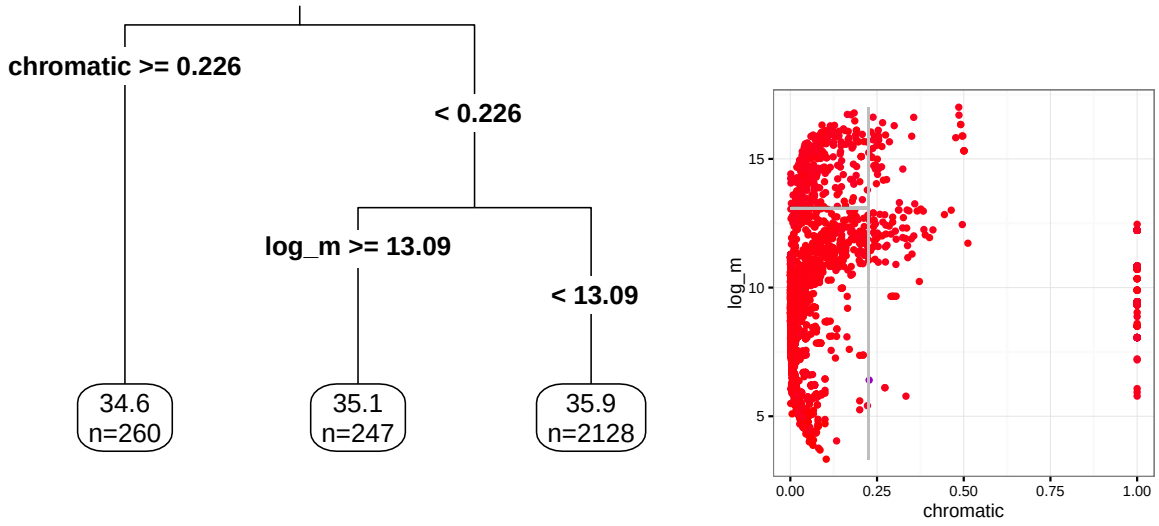

Figure 29: A CART model identifying instances on which MER99MU performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

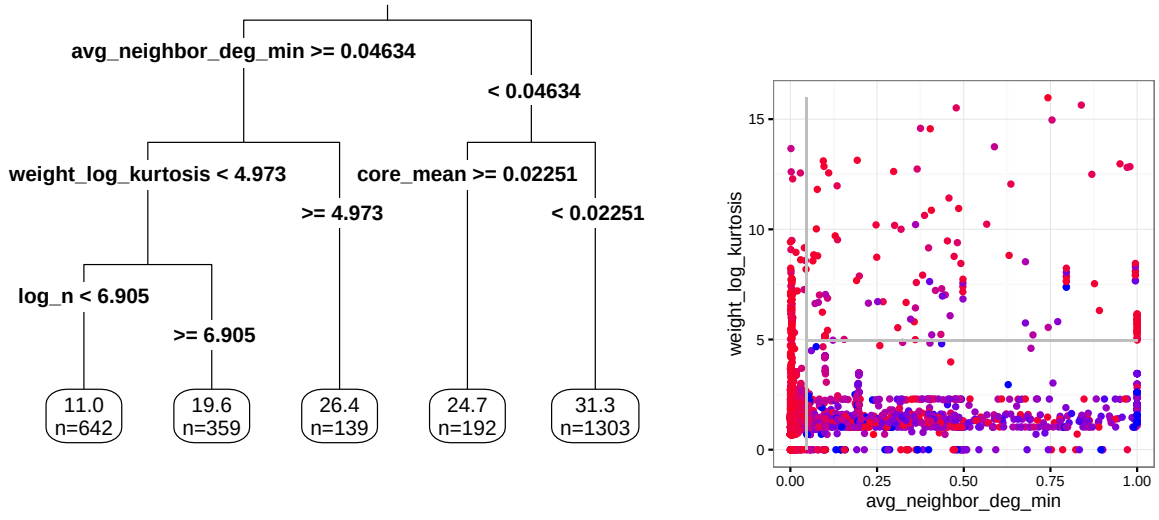

Figure 30: A CART model identifying instances on which PAL04MT performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

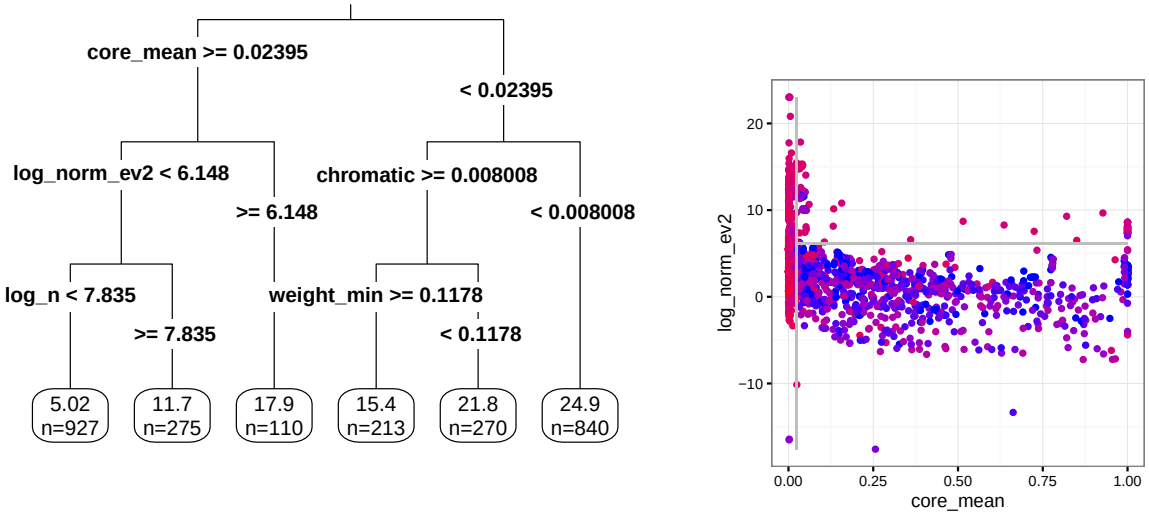

Figure 31: A CART model identifying instances on which PAL04T1 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

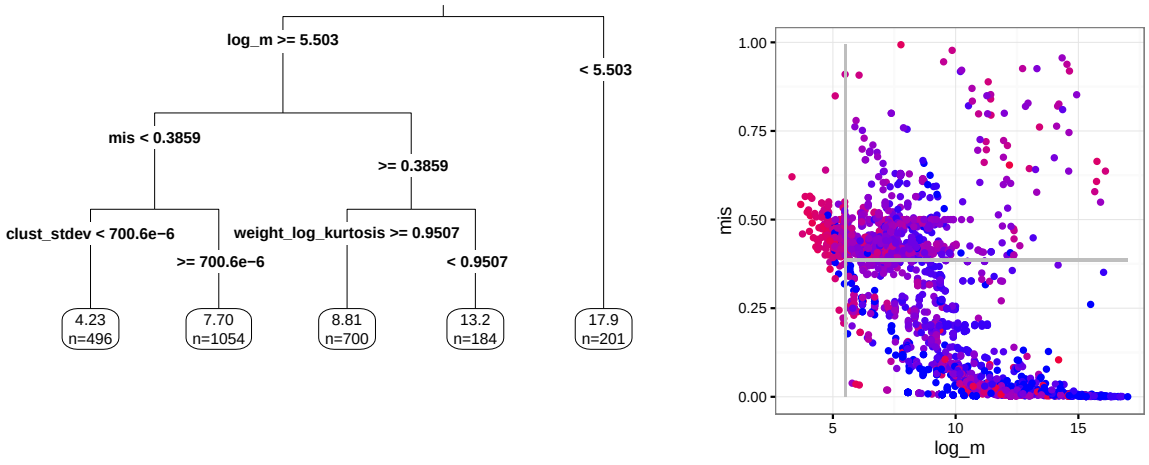

Figure 32: A CART model identifying instances on which PAL04T2 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

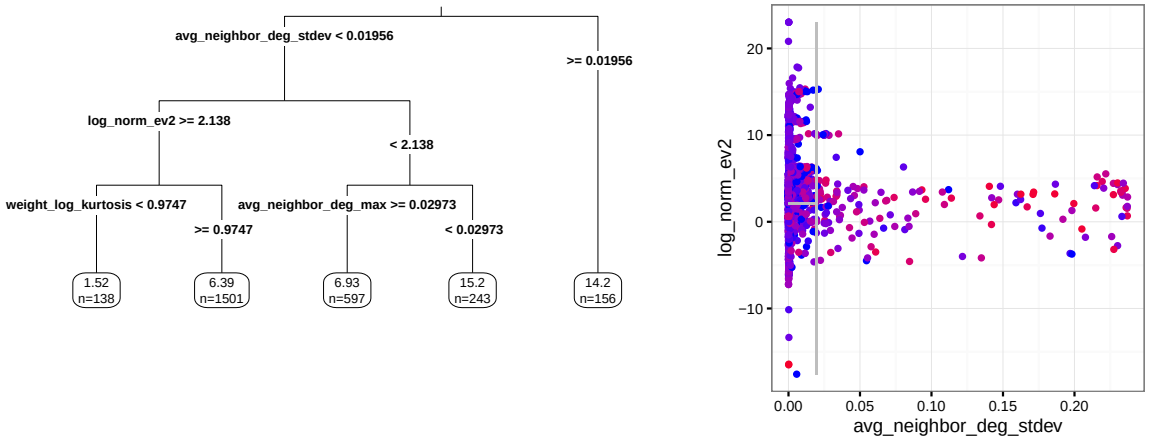

Figure 33: A CART model identifying instances on which PAL04T3 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

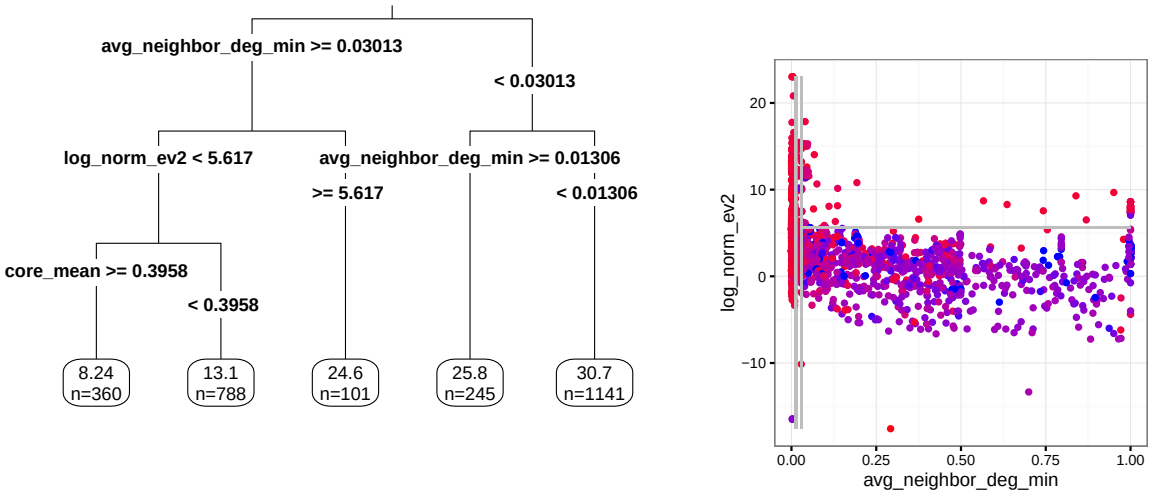

Figure 34: A CART model identifying instances on which PAL04T4 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

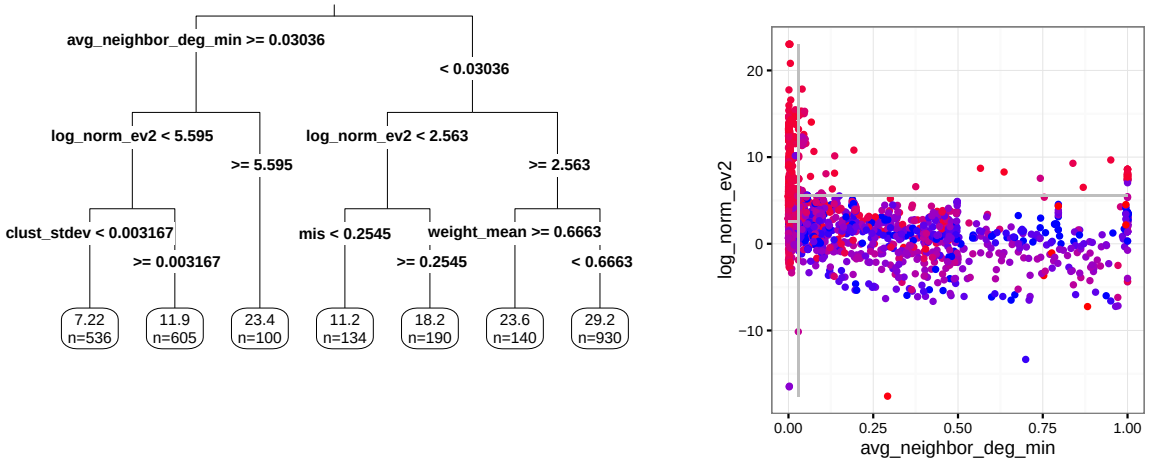

Figure 35: A CART model identifying instances on which PAL04T5 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

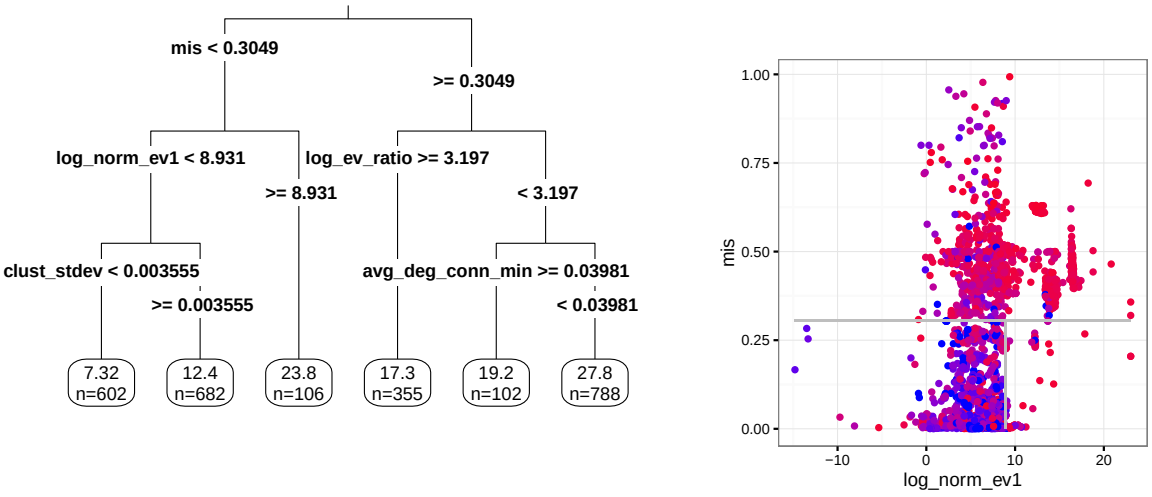

Figure 36: A CART model identifying instances on which PAL06 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).

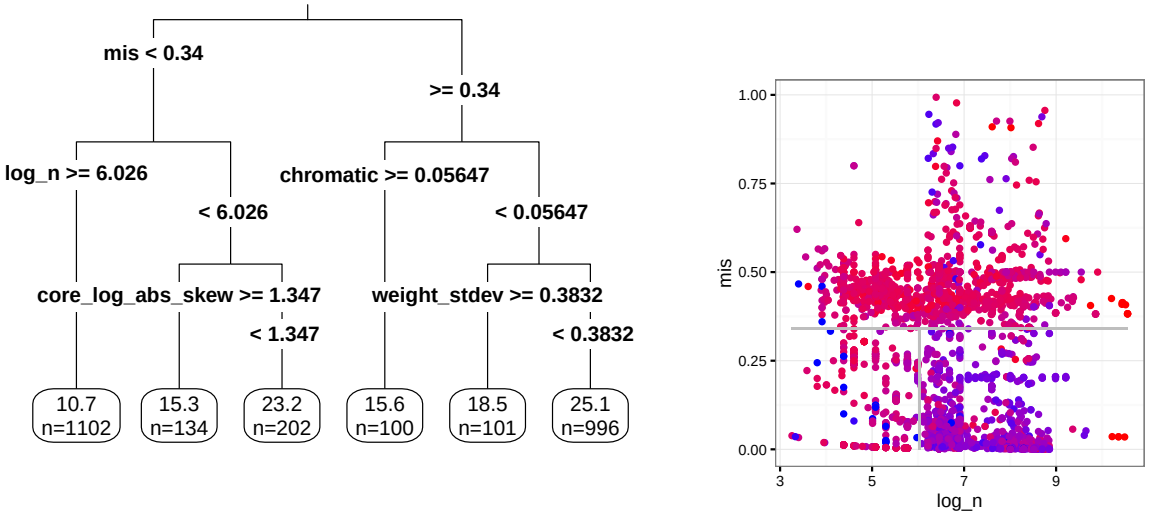

Figure 37: A CART model identifying instances on which PAR08 performs particularly well or poorly. Blue indicates the heuristic performed well (rank near 1) and red indicates the heuristic performed poorly (rank near 37).