	WHERE		CART		Random Forest	
Data set	default	Tuned	default	Tuned	default	Tuned
antV0	30	**89**	27	**89**	28	**89**
antV1	32	**74**	36	**74**	41	**74**
antV2	78	52	52	56	64	**100**
camelV0	**83**	**83**	26	34	50	78
camelV1	22	81	24	**83**	25	31
ivyV0	16	**89**	17	25	16	21
jeditV0	35	48	49	**61**	44	50
jeditV1	24	**87**	28	62	28	37
jeditV2	2	**98**	3	18	5	5
log4jV0	94	**100**	97	**100**	99	**100**
luceneV0	61	76	67	70	67	**77**
poiV0	74	74	**79**	75	77	78
poiV1	**100**	75	73	89	82	**100**
synapseV0	66	66	71	**100**	65	**100**
velocityV0	34	**43**	34	39	34	42
xercesV0	13	**85**	14	47	15	14
xercesV1	**56**	26	49	26	49	26

Figure 1: Exp A: Precision results (best results shown in bold).

	WHERE		CART		Random Forest	
Data set	default	Tuned	default	Tuned	default	Tuned
antV0	39	20	32	40	**41**	21
antV1	11	5	38	49	37	**55**
antV2	0	3	44	**48**	47	46
camelV0	0	**91**	9	28	4	31
camelV1	35	35	31	33	**37**	33
ivyV0	28	32	28	30	27	**33**
jeditV0	50	**57**	56	**57**	56	**57**
jeditV1	37	38	36	45	43	**49**
jeditV2	4	6	5	**9**	**9**	8
log4jV0	**61**	54	47	46	57	45
luceneV0	70	74	56	74	70	**75**
poiV0	82	71	79	62	**83**	73
poiV1	5	**78**	21	**78**	19	**78**
synapseV0	0	2	40	55	41	**56**
velocityV0	51	51	49	**53**	51	51
xercesV0	22	23	21	**27**	24	20
xercesV1	23	4	18	**47**	18	40

Figure 2: Exp A: F-value results (best results shown in bold).

	Precision		F		SUM		
Features	default	tuned	default	tuned	default	tuned	
max_cc							
noc							
ca							
cbo		1				1	
moa		1				1	
ce		2		1		3	
avg_cc		2		2		4	
npm	1	2	1	1	2	3	
lcom	1	2	1	1	2	3	
amc	4	2	4	2	8	4	
cbm	5	2	5	3	10	5	
rfc	3	6	3	8	6	14	
wmc	5	4	5	9	10	13	
dit	8	3	7	7	15	10	
ic	8	3	8	6	16	9	
lcom3	8	6	8	8	16	14	
cam	9	7	9	7	18	14	
loc	9	5	9	11	18	16	
dam	13	6	13	11	26	17	
mfa	16	9	16	9	32	18	

Figure 3: Exp A: Counts of features selected by different goals.Given that we are processing 17 data sets, the maximum counts for any one cell in the "precision" or "F" column is 17.

Datasets	Tuned_Where	Naive_Where	Tuned_CART	Naive_CART	Tuned_RanFst	Naive_RanFst
ant	50/ 90.18	1.57	60/ 4.36	0.09	50/ 8.12	0.18
antV1	50/ 174.67	2.90	50/ 6.35	0.10	50/ 11.77	0.27
antV2	50/ 403.63	6.92	70/ 9.71	0.16	60/ 13.28	0.35
camel	50/ 537.53	8.60	50/ 9.14	0.18	50/ 13.73	0.31
camelV1	60/ 1640.54	24.57	60/ 17.06	0.24	70/ 31.53	0.73
ivy	70/ 77.75	1.02	60/ 3.86	0.06	50/ 8.00	0.17
jedit	80/ 472.57	5.49	60/ 6.30	0.09	70/ 13.01	0.30
jeditV1	60/ 489.45	6.82	70/ 7.61	0.11	60/ 12.87	0.32
jeditV2	50/ 435.43	7.21	50/ 6.46	0.12	90/ 20.34	0.37
log4j	70/ 113.73	1.36	70/ 3.25	0.05	60/ 7.07	0.16
lucene	70/ 224.39	2.70	50/ 4.07	0.08	50/ 8.87	0.26
poi	60/ 261.06	4.00	60/ 6.23	0.10	50/ 10.57	0.29
poiV1	80/ 607.85	7.18	60/ 7.69	0.13	50/ 11.39	0.29
synapse	50/ 116.04	1.87	60/ 4.07	0.05	70/ 9.74	0.16
velocity	60/ 195.27	2.75	60/ 4.49	0.06	80/ 12.15	0.21
xerces	60/ 143.69	2.17	70/ 7.26	0.09	60/ 10.28	0.23
xercesV1	50/ 794.50	13.37	50/ 8.24	0.15	60/ 14.54	0.38

Figure 4: Time (in seconds) spent on different models over the objective of prec

Datasets	Tuned_Where	Naive_Where	Tuned_CART	Naive_CART	Tuned_RanFst	Naive_RanFst
ant	50/ 94.08	1.71	60/ 4.55	0.08	60/ 10.79	0.21
antV1	60/ 193.74	3.02	70/ 7.77	0.09	60/ 12.30	0.25
antV2	80/ 643.94	7.59	60/ 8.38	0.15	70/ 16.99	0.41
camel	60/ 662.56	9.97	60/ 13.19	0.23	80/ 26.11	0.32
camelV1	60/ 1800.64	24.25	50/ 15.02	0.28	50/ 28.52	0.78
ivy	60/ 69.95	1.03	50/ 3.35	0.08	70/ 9.40	0.18
jedit	90/ 553.80	5.58	50/ 5.58	0.09	60/ 15.08	0.33
jeditV1	60/ 519.75	8.76	50/ 7.43	0.13	60/ 18.13	0.41
jeditV2	70/ 621.32	8.98	50/ 9.71	0.15	60/ 17.38	0.63
log4j	70/ 125.29	1.73	50/ 2.90	0.06	60/ 8.76	0.19
lucene	50/ 221.99	3.52	50/ 5.20	0.10	50/ 10.09	0.33
poi	60/ 327.48	5.13	50/ 6.56	0.11	50/ 12.88	0.36
poiV1	50/ 523.85	8.95	80/ 12.26	0.14	60/ 19.56	0.35
synapse	70/ 148.23	1.91	60/ 3.96	0.06	60/ 8.19	0.16
velocity	50/ 156.51	2.75	60/ 4.27	0.06	50/ 7.70	0.22
xerces	60/ 142.83	2.01	70/ 7.15	0.08	60/ 9.61	0.20
xercesV1	50/ 751.92	12.98	60/ 9.28	0.16	50/ 12.69	0.38

Figure 5: Time (in seconds) spent on different models over the objective of F

Learner Name	Parameters	Default	antV0	antV1	antV2	camelV0	camelV1	ivyV0	jeditV0	jeditV1	jeditV2	log4jV0	luceneV0	poiV0	poiV1	synapseV0	velocityV0	xercesV0	xercesV1
Where based Learner	threshold	0.5	0.98	0.98	0.43	0.24	0.64	1	1	0.98	0.98	1	1	0.87	0.59	0.98	1	0.98	0.98
	infoPrune	0.33	0.05	0.05	0.71	0.54	0.45	0.41	0.3	0.05	0.05	0.54	0.84	0.01	1	0.05	0.68	0.43	0.05
	min_sample_size	4	7	7	9	8	6	10	1	5	7	8	7	9	3	7	7	1	7
	min_Size	0.5	0.51	0.51	0.59	0.46	0.13	0.38	0.66	0.27	0.51	0.46	0.47	0.77	0.48	0.51	0.66	0.22	0.51
	wriggle	0.2	0.6	0.6	0.83	0.52	0.19	0.01	0.26	0.6	0.6	0.52	0.19	0.83	0.01	0.6	0.26	0.55	0.6
	depthMin	2	1	1	2	3	5	2	3	3	1	1	2	4	2	1	3	2	1
	depthMax	10	8	8	13	19	19	18	7	8	8	19	1	19	18	8	11	18	8
	wherePrune	False	False	False	False	False	True	True	False	False	False	False	False	True	True	False	False	True	False
	treePrune	True	False	False	False	True	True	True	False	False	False	True	True	False	False	False	False	True	False
CART	threshold	0.5	0.69	0.99	1	0.3	0.83	1	0.99	0.58	0.72	1	0.71	0.46	0.72	1	0.85	1	0.64
	max_feature	None	0.01	0.58	0.65	0.66	0.73	0.67	0.56	0.01	0.97	0.54	0.52	0.32	0.01	0.74	0.73	0.01	0.1
	min_samples_split	2	7	16	18	5	11	6	15	6	17	4	16	12	5	14	11	4	10
	min_samples_leaf	1	13	14	10	4	3	15	16	9	6	7	6	1	4	6	7	11	7
	max_depth	None	14	1	41	34	1	22	29	1	1	1	14	19	8	40	4	1	1
Random Forests	threshold	0.5	0.84	0.9	0.83	0.33	1	0.99	0.91	1	1	0.83	0.98	0.9	0.86	0.83	1	1	1
	max_feature	None	0.61	0.13	0.89	0.37	0.01	0.98	0.52	0.75	0.35	0.01	0.98	0.84	0.73	0.01	0.48	0.51	0.01
	max_leaf_nodes	None	37	35	38	21	36	45	10	38	10	30	20	43	11	13	15	39	10
	min_samples_split	2	8	16	17	13	14	2	3	2	2	18	19	9	4	4	9	20	1
	min_samples_leaf	1	19	5	2	4	2	4	7	17	7	16	12	2	3	2	2	2	3
	n_estimators	100	138	112	77	74	125	130	107	85	96	111	103	82	59	149	150	63	58

Figure 6: Parameters tuned on different models over the objective of prec

Learner Name	Parameters	Default	antV0	antV1	antV2	camelV0	camelV1	ivyV0	jeditV0	jeditV1	jeditV2	log4jV0	luceneV0	poiV0	poiV1	synapseV0	velocityV0	xercesV0	xercesV1
Where based Learner	threshold	0.5	0.04	0.44	0.44	0.98	0.65	0.77	1	0.65	0.98	0.44	0.44	0.87	0.04	0.77	0.24	0.44	0.77
	infoPrune	0.33	0.51	0.68	0.88	0.47	0.07	0.31	0.48	0.68	0.57	0.12	0.68	0.01	0.51	0.14	0.54	0.68	0.14
	min_sample_size	4	6	4	6	1	6	8	8	4	6	7	4	9	6	2	8	4	8
	min_Size	0.5	0.18	0.4	0.56	0.51	0.65	0.59	0.97	0.4	0.51	0.8	0.4	0.77	0.18	0.62	0.46	0.4	0.66
	wriggle	0.2	0.25	0.29	0.76	0.6	0.63	0.26	1	0.51	0.17	0.36	0.51	0.83	0.25	0.5	0.52	0.29	0.26
	depthMin	2	3	3	3	1	5	3	2	3	5	5	3	4	3	3	3	3	3
	depthMax	10	16	15	15	8	19	10	7	15	5	15	15	19	16	6	19	15	10
	wherePrune	False	False	True	True	True	True	True	True	False	False	True	True	True	False	True	False	False	True
	treePrune	True	False	True	True	False	False	False	False	False	True	True	True	False	False	False	True	True	False
CART	threshold	0.5	0.34	0.25	0.01	0.01	0.73	0.53	0.92	0.8	0.74	0.54	0.03	0.91	0.01	0.01	0.55	1	0.01
	max_feature	None	0.01	0.01	0.29	0.01	0.46	0.75	0.79	0.74	0.41	0.81	0.61	0.72	0.01	0.01	0.01	0.25	0.18
	min_samples_split	2	18	20	12	2	15	11	2	18	13	9	17	16	10	4	8	3	15
	min_samples_leaf	1	19	16	15	17	1	1	13	10	4	3	7	5	20	7	8	1	6
	max_depth	None	12	2	15	1	41	20	44	15	13	5	23	14	1	5	17	47	13
Random Forests	threshold	0.5	0.01	0.35	0.3	0.01	0.9	0.97	0.63	1	0.73	0.68	0.01	1.0	0.01	0.07	0.22	1	0.82
	max_feature	None	0.63	0.17	0.01	0.01	0.88	0.74	0.76	0.73	0.01	0.03	0.39	0.02	0.01	0.56	0.36	0.51	0.89
	max_leaf_nodes	None	40	33	46	22	11	16	38	34	30	31	12	49	25	47	15	39	24
	min_samples_split	2	10	16	20	1	1	1	1	4	20	19	11	14	2	17	19	20	19
	min_samples_leaf	1	4	15	9	13	18	11	3	16	17	6	10	7	19	13	11	2	14
	n_estimators	100	120	73	75	130	97	144	125	97	80	111	96	101	50	67	74	63	66

Figure 7: Parameters tuned on different models over the objective of F