

Appendix A. Base Learner statistics

#	pruning	representation	lemmas	n-grams	stopwords	LSA [dims]	undersampling	model	F1 @ val
0	X	TfIdf	✓	3	X	X	✓	LogisticRegression	55.73%
1	window=0	TfIdf	✓	1	✓	200	✓	SVM(RBF)	54.29%
2	window=5	BinaryCounts	✓	1	✓	200	✓	LogisticRegression	53.57%
3	SDP	BinaryCounts	✓	1	✓	200	✓	RandomForest	53.33%
4	window=0	BinaryCounts	✓	3	X	200	✓	RandomForest	53.85%
0	X	BinaryCounts	✓	3	X	X	✓	SVM(RBF)	56.84%
1	SDP	TfIdf	✓	3	X	200	✓	SVM(linear)	56.79%
2	window=5	TfIdf	✓	3	X	X	✓	SVM(linear)	60.80%
3	window=5	TfIdf	✓	1	✓	200	✓	SVM(linear)	52.87%
4	window=5	TfIdf	✓	3	X	X	✓	RandomForest	53.58%
5	—	—	—	—	—	—	—	LSTM	44.43%
6	window=0	TfIdf	✓	3	X	200	✓	SVM(linear)	57.87%
7	X	TfIdf	✓	1	✓	200	✓	SVM(RBF)	50.30%
8	SDP	BinaryCounts	✓	1	✓	200	✓	LogisticRegression	51.67%
9	window=0	TfIdf	✓	1	✓	200	✓	SVM(RBF)	54.29%
0	window=5	TfIdf	✓	3	X	X	✓	SVM(linear)	60.80%
1	window=0	TfIdf	✓	3	X	X	✓	SVM(RBF)	52.02%
2	X	BinaryCounts	✓	3	X	X	✓	SVM(RBF)	56.84%
3	window=0	BinaryCounts	✓	3	X	200	✓	LogisticRegression	57.21%
4	SDP	TfIdf	✓	3	X	200	✓	LogisticRegression	56.52%
5	window=5	TfIdf	✓	1	✓	200	✓	SVM(linear)	52.87%
6	window=5	TfIdf	✓	3	X	X	✓	RandomForest	53.58%
7	—	—	—	—	—	—	—	LSTM	44.43%
8	SDP	TfIdf	✓	3	X	200	✓	RandomForest	52.35%
9	X	TfIdf	✓	1	✓	200	✓	SVM(RBF)	51.43%
10	window=0	TfIdf	✓	1	✓	200	✓	SVM(RBF)	54.29%
11	window=5	BinaryCounts	✓	1	✓	X	✓	RandomForest	51.29%
12	SDP	TfIdf	✓	1	✓	200	✓	SVM(RBF)	51.63%
0	SDP	TfIdf	✓	1	✓	200	✓	SVM(RBF)	51.63%
1	window=5	TfIdf	✓	1	✓	X	✓	SVM(RBF)	53.12%
2	window=0	TfIdf	✓	3	X	X	✓	SVM(RBF)	52.02%
3	window=5	BinaryCounts	✓	3	X	X	✓	SVM(RBF)	61.30%
4	window=0	BinaryCounts	✓	3	X	200	✓	LogisticRegression	57.21%
5	window=5	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	54.18%
6	SDP	TfIdf	✓	3	X	200	✓	LogisticRegression	56.52%
7	—	—	—	—	—	—	—	LSTM	44.43%

8	SDP	TfIdf	✓	1	✓	✗	✓	RandomForest	49.78%
9	✗	BinaryCounts	✓	3	✗	✗	✓	SVM(RBF)	56.84%
10	window=0	TfIdf	✓	1	✓	200	✓	SVM(RBF)	54.29%
11	✗	TfIdf	✓	1	✓	200	✓	LogisticRegression	50.95%
12	SDP	TfIdf	✓	3	✗	✗	✓	RandomForest	49.69%
13	window=5	TfIdf	✓	3	✗	200	✓	RandomForest	54.76%
14	✗	BinaryCounts	✓	3	✗	200	✓	RandomForest	50.67%
0	✗	BinaryCounts	✓	3	✗	200	✓	SVM(linear)	53.34%
1	window=5	TfIdf	✓	1	✓	200	✓	SVM(RBF)	52.90%
2	SDP	TfIdf	✓	1	✓	✗	✓	RandomForest	49.78%
3	SDP	TfIdf	✓	1	✓	200	✓	SVM(RBF)	51.63%
4	✗	TfIdf	✓	1	✓	200	✓	LogisticRegression	50.95%
5	✗	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	56.84%
6	window=5	TfIdf	✓	3	✗	200	✓	SVM(RBF)	50.34%
7	window=0	TfIdf	✓	3	✗	200	✓	SVM(RBF)	55.78%
8	window=0	BinaryCounts	✓	3	✗	200	✓	SVM(RBF)	57.89%
9	window=0	TfIdf	✓	1	✓	200	✓	RandomForest	53.97%
10	SDP	TfIdf	✓	3	✗	200	✓	LogisticRegression	56.52%
11	—	—	—	—	—	—	—	LSTM	44.43%
12	window=5	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	54.18%
13	window=0	TfIdf	✓	3	✗	200	✓	RandomForest	50.85%
14	window=5	BinaryCounts	✓	3	✗	200	✓	SVM(linear)	57.45%
15	SDP	BinaryCounts	✓	3	✗	✗	✓	SVM(linear)	54.43%
16	window=0	TfIdf	✓	1	✓	200	✓	SVM(RBF)	54.29%
17	window=5	BinaryCounts	✓	1	✓	✗	✓	RandomForest	51.29%
18	window=5	BinaryCounts	✓	3	✗	✗	✓	SVM(RBF)	61.30%
19	SDP	TfIdf	✓	3	✗	✗	✓	RandomForest	49.69%

Table 1: Base Learner selection (CPR task)

#	pruning	representation	lemmas	n-grams	stopwords	LSA [dims]	undersampling	model	F1 @ val
0	window=5	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	51.88%
1	window=0	TfIdf	✓	1	✓	✗	✓	SVM(linear)	49.94%
2	SDP	TfIdf	✓	1	✗	200	✓	SVM(linear)	56.21%
3	SDP	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	49.70%
4	window=5	TfIdf	✓	1	✗	✗	✓	LogisticRegression	51.51%
0	window=5	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	51.88%
1	SDP	TfIdf	✓	1	✗	200	✓	SVM(linear)	56.21%
2	window=5	BinaryCounts	✓	1	✗	200	✓	SVM(linear)	50.46%

3	SDP	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	49.70%
4	window=5	TfIdf	✓	1	✗	✗	✓	LogisticRegression	51.51%
5	SDP	TfIdf	✓	1	✗	✗	✓	RandomForest	49.84%
0	SDP	BinaryCounts	✓	1	✓	✗	✓	SVM(linear)	49.85%
1	window=0	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	52.62%
2	window=5	BinaryCounts	✓	1	✗	200	✓	SVM(RBF)	52.08%
3	SDP	TfIdf	✓	1	✗	200	✓	SVM(linear)	56.21%
4	SDP	BinaryCounts	✓	1	✓	✗	✓	RandomForest	50.00%
5	✗	BinaryCounts	✓	1	✗	200	✓	SVM(RBF)	50.28%
6	✗	TfIdf	✓	1	✗	✗	✓	RandomForest	50.36%
7	SDP	TfIdf	✓	1	✓	200	✓	LogisticRegression	52.47%
8	✗	TfIdf	✓	1	✗	200	✓	SVM(RBF)	49.92%
9	window=0	BinaryCounts	✓	1	✓	✗	✓	LogisticRegression	50.56%
0	window=5	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	51.88%
1	SDP	BinaryCounts	✓	1	✓	✗	✓	SVM(linear)	49.85%
2	window=0	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	52.62%
3	✗	TfIdf	✓	1	✗	✗	✓	RandomForest	50.36%
4	SDP	TfIdf	✓	1	✗	200	✓	SVM(linear)	56.21%
5	SDP	BinaryCounts	✓	1	✓	✗	✓	RandomForest	50.00%
6	window=5	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	51.20%
7	✗	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	49.88%
8	✗	BinaryCounts	✓	1	✗	200	✓	SVM(RBF)	50.28%
9	SDP	TfIdf	✓	1	✓	200	✓	LogisticRegression	52.47%
10	window=0	TfIdf	✓	1	✗	200	✓	RandomForest	50.00%
11	✗	BinaryCounts	✓	1	✓	200	✓	RandomForest	50.38%
12	window=0	BinaryCounts	✓	1	✓	✗	✓	LogisticRegression	50.56%
0	window=5	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	51.88%
1	window=0	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	52.62%
2	✗	TfIdf	✓	1	✗	✗	✓	RandomForest	50.36%
3	window=0	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	52.04%
4	SDP	TfIdf	✓	1	✗	200	✓	SVM(linear)	56.21%
5	window=0	BinaryCounts	✓	1	✓	200	✓	LogisticRegression	50.50%
6	window=5	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	51.20%
7	✗	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	49.88%
8	SDP	BinaryCounts	✓	1	✓	✗	✓	SVM(linear)	49.85%
9	SDP	TfIdf	✓	1	✓	200	✓	LogisticRegression	52.47%
10	window=0	TfIdf	✓	1	✗	200	✓	RandomForest	50.00%
11	window=5	BinaryCounts	✓	1	✗	200	✓	SVM(linear)	50.46%
12	✗	BinaryCounts	✓	1	✓	200	✓	RandomForest	50.38%
13	window=5	BinaryCounts	✓	1	✗	✗	✓	RandomForest	49.85%
14	SDP	TfIdf	✓	1	✗	✗	✓	RandomForest	49.84%
0	window=5	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	51.88%
1	window=0	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	52.62%

2	window=0	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	52.04%
3	✗	TfIdf	✓	1	✗	✗	✓	RandomForest	50.36%
4	window=5	BinaryCounts	✓	1	✗	✗	✓	RandomForest	49.85%
5	SDP	TfIdf	✓	1	✓	✗	✓	SVM(linear)	49.86%
6	window=0	BinaryCounts	✓	1	✓	200	✓	LogisticRegression	50.50%
7	✗	BinaryCounts	✓	1	✓	200	✓	RandomForest	50.38%
8	window=5	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	51.20%
9	✗	TfIdf	✓	1	✗	✗	✓	SVM(RBF)	49.88%
10	SDP	TfIdf	✓	1	✓	200	✓	LogisticRegression	52.47%
11	SDP	TfIdf	✓	1	✗	200	✓	SVM(RBF)	54.34%
12	window=5	BinaryCounts	✓	1	✗	200	✓	RandomForest	52.92%
13	window=0	TfIdf	✓	1	✗	200	✓	RandomForest	50.00%
14	SDP	BinaryCounts	✓	1	✗	200	✓	SVM(RBF)	54.55%
15	window=5	BinaryCounts	✓	1	✗	200	✓	SVM(linear)	50.46%
16	SDP	BinaryCounts	✓	1	✓	200	✓	SVM(RBF)	49.70%
17	SDP	TfIdf	✓	1	✗	✗	✓	RandomForest	49.84%
18	SDP	BinaryCounts	✓	1	✗	✗	✓	SVM(linear)	56.95%
19	window=5	BinaryCounts	✓	1	✗	✗	✓	SVM(RBF)	52.15%

Table 2: Base Learner selection (CID task)