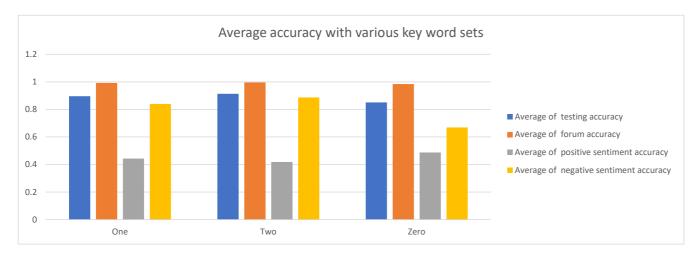
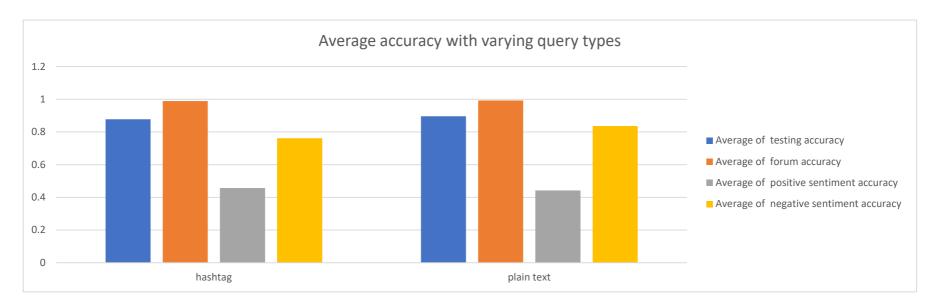
key words	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
One	0.896009912	0.992675281	0.44309375	0.839975
Two	0.913374717	0.995841863	0.4183125	0.88674375
Zero	0.851387414	0.984359007	0.48726875	0.66935



Key word set 2 appears to have the highest testing, forum and negative One sentiment accuracy. The reduction in positive sentiment accuracy is disturbing but will most likely be negligible if a model with appropriate Zero pruning methods is chosen

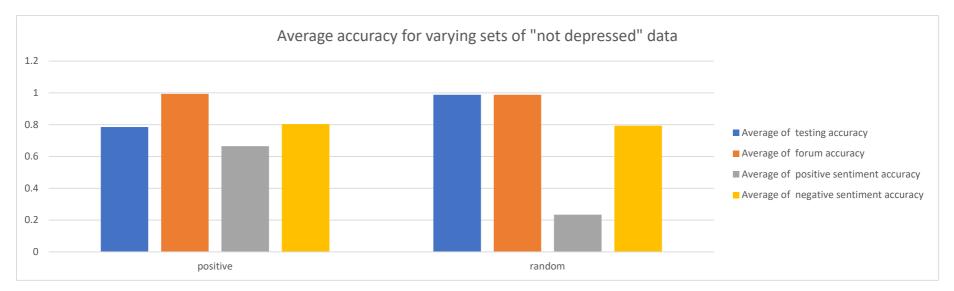
"sad", "unhappy", "stressed", "anxious", "miserable"
"anxiety", "stress", "despondent", "suicidal", "depressed"
"depression", "suicide", "mental illness", "severe anxiety", "depression medication"

query type	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
hashtag	0.877927483	0.989018253	0.457308333	0.761520833
plain text	0.895920545	0.992899181	0.441808333	0.835858333



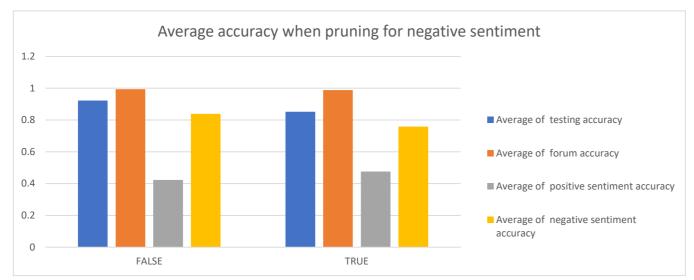
Plain text shows higher accuracy in almost every field except positive sentiment accuracy but even then only at a negligible discrepency with hashtag queries

not depressed set type	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
positive	0.785276282	0.993645513	0.664620833	0.803925
random	0.988571746	0.988271921	0.234495833	0.793454167



Positive shows higher average accuracy in all but testing accuracy to be expected given the harsh contrast of the resulting dataset, but a model with higher testing accuracy will most likely be chosen. It is to be noted that positive datasets score have less errors in identifying not depressed posts

negative sentiment filter	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
FALSE	0.922198831	0.993560218	0.423104167	0.839208333
TRUE	0.851649197	0.988357216	0.4760125	0.758170833



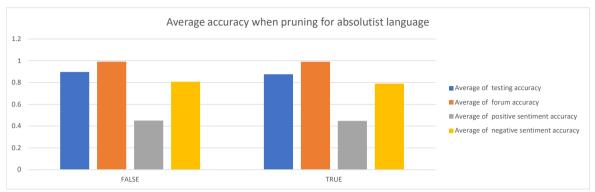
Negative sentiment filtering has lower testing, forum and negative sentiment accuracy. However, the impact of negative sentiment pruning on reducing errors in false-positives is not to be understated

personal pronouns	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
FALSE	0.906520376	0.992024906	0.452554167	0.8131
TRUE	0.867327652	0.989892528	0.4465625	0.784279167



As to be expected in all pruning methods that reduce dataset sizes, personal pronoun pruning does see lower accuracy in almost every field. It appears that the positive affect of personal pronouns on selecting accurate data is negligible.

absolutist language	Average of testing accuracy	Average of forum accuracy	Average of positive sentiment accuracy	Average of negative sentiment accuracy
FALSE	0.897645066	0.991491812	0.450908333	0.807533333
TRUE	0.876202962	0.990425623	0.448208333	0.789845833



Much like pruning for personal pronouns, pruning for absolutist language does reduce accuracy. However, like personal pronouns, the reduction is minimal and the likely increase in real world accuracy makes this form of pruning desirable in the dataset

id	key words	query type	not depressed set type	negative sentiment filter	nersonal nronouns	absolutist language	testing accuracy	positive sentiment ao	negative contiment a	fonim accuracy
PM I		hashtag	random	FALSE	FALSE	FALSE	100%	31%	63%	979
	1 0		positive	FALSE	FALSE	FALSE	90%	85%	64%	999
	2 0		random	FALSE	FALSE	TRUE	99%	37%	58%	979
1			positive	FALSE	FALSE	TRUE	83%	82%	62%	999
		hashtag	random	FALSE	TRUE	FALSE	99%	38%	58%	989
	5 0	hashtag	positive	FALSE	TRUE	FALSE	74%	74%	58%	989
		hashtag	random	FALSE	TRUE	TRUE	98%	41%	57%	989
		hashtag	positive	FALSE	TRUE	TRUE	63%	49%	59%	989
1	8 0	hashtar	random	TRUE	FALSE	FALSE	96%	46%	55%	985
1		hashtag hashtag	positive	TRUE TRUE	FALSE FALSE	FALSE TRUE	66% 95%	44% 47%	60% 55%	98:
1	0	hashtag hashtag		TRUE	FALSE FALSE	TRUE		47%	55% 60%	98:
1			positive	TRUE	TRUE	FALSE	62% 94%	42%	55%	
1	3 0	hashtag	positive	TRUE	TRUE	FALSE	60%	41%	59%	981
1	4 0	hashtag	random	TRUE	TRUE	TRUE	94%	48%	55%	98
1	5 0		positive	TRUE	TRUE	TRUE	62%	40%	59%	98
1	6 0	plain text	random	FALSE	FALSE	FALSE	100%	10%	88%	100
1	7 0	plain text	positive	FALSE	FALSE	FALSE	93%	79%	82%	100
1	8 0	plain text	random	FALSE	FALSE	TRUE	100%	15%	80%	99
1	9 0	plain text	positive	FALSE	FALSE	TRUE	86%	76%	83%	100
2	0	plain text	random	FALSE	TRUE	FALSE	99%	18%	78%	99
2	1 0	plain text	positive	FALSE	TRUE	FALSE	78%	71%	83%	100
2		plain text	random	FALSE	TRUE	TRUE	99%	24%	74%	99
2		plain text	positive	FALSE	TRUE	TRUE	68%	63%	77%	99
2		plain text	random	TRUE	FALSE	FALSE	99%	39%	69%	99
2		plain text	positive	TRUE	FALSE	FALSE	70%	63%	73%	99
2		plain text	random	TRUE	FALSE	TRUE	98%	41%	68%	98
2		plain text	positive	TRUE	FALSE	TRUE	69%	61%	73%	99
2		plain text	random	TRUE	TRUE	FALSE	99%	42%	67%	98
2		plain text	positive	TRUE	TRUE	FALSE	67%	61%	72%	99
3	0	plain text	random	TRUE TRUE	TRUE TRUE	TRUE	99%	43% 61%	66% 72%	98
3		plain text hashtag	positive	FALSE	FALSE	FALSE	100%		72% 94%	
		hashtag hashtag	random positive	FALSE FALSE	FALSE	FALSE FALSE		5% 76%	94% 88%	100
3	3 1 4 1		positive	FALSE FALSE	FALSE FALSE	TRUE	95% 99%	76% 6M	88% 93%	100
3	5 1		positive	FALSE	FALSE	TRUE	90%	73%	88%	100
3		hashtag	random	FALSE	TRUE	FALSE	99%	11%	87%	99
3			positive	FALSE	TRUE	FALSE	83%	68%	88%	100
3		hashtag	random	FALSE	TRUE	TRUE	99%	18%	80%	99
3			positive	FALSE	TRUE	TRUE	73%	64%	86%	100
4		hashtag	random	TRUE	FALSE	FALSE	99%	32%	77%	99
4	1 1	hashtag	positive	TRUE	FALSE	FALSE	74%	68%	79%	99
4	2 1	hashtag	random	TRUE	FALSE	TRUE	99%	33%	75%	98
4		hashtag	positive	TRUE	FALSE	TRUE	72%	67%	78%	99
4		hashtag	random	TRUE	TRUE	FALSE	99%	34%	75%	99
4			positive	TRUE	TRUE	FALSE	71%	67%	76%	99
4	6 1	hashtar	random	TRUE	TRUE	TRUE	99%	36%	73%	99
4		hashtag	positive	TRUE	TRUE	TRUE	73%	68%	75%	99
4	8 1		random	FALSE	FALSE	FALSE FALSE	100%	4%	96%	100
4			positive	FALSE	FALSE			76%	89%	
5		plain text plain text	random positive	FALSE FALSE	FALSE FALSE	TRUE	100%	5%	95% 90%	100
5	1 1	plain text	random	FALSE	TRUE	FALSE	92% 99%	73% 9%	90%	100 99
5.		plain text	positive	FALSE	TRUE	FALSE	86%	69%	90%	100
5	4 1	plain text	random	FALSE	TRUE	TRUE	99%	14%	86%	99
5		plain text	positive	FALSE	TRUE	TRUE	77%	63%	89%	100
5			random	TRUE	FALSE	FALSE	99%	25%	83%	99
5		plain text	positive	TRUE	FALSE	FALSE	76%	68%	86%	100
5			random	TRUE	FALSE	TRUE	99%	27%	81%	99
5	9 1	plain text	positive	TRUE	FALSE	TRUE	74%	67%	85%	100
6	0 1	plain text	random	TRUE	TRUE	FALSE	99%	28%	79%	99
6		plain text	positive	TRUE	TRUE	FALSE	75%	67%	81%	99
6			random	TRUE	TRUE	TRUE	99%	31%	78%	99
6		plain text	positive	TRUE	TRUE	TRUE	73%	66%	80%	99
6		hashtag	random	FALSE	FALSE	FALSE	100%	4%	96%	100
6		hashtag	positive	FALSE	FALSE	FALSE	95%	74%	91%	100
6	6 2	hashtar	random	FALSE	FALSE	TRUE	100%	4%	96%	100
6	2	hashtag	positive	FALSE	FALSE	TRUE	93%	73%	90%	100
6	2	hashtag	random	FALSE FALSE	TRUE TRUE	FALSE FALSE	99% 89%	5%	95% 91%	100
6	2	hashtag hashtag	positive random	FALSE FALSE	TRUE	FALSE TRUE	89% 99%	69% 12%	91% 87%	100
7	. 2	hashtag hashtag	random	FALSE FALSE	TRUE	TRUE		12%	87% 89%	100
7.	2	hashtar hashtar	random	FALSE	FALSE	FALSE	79% 99%	63% 24%	89%	100
7.		hashtag hashtag	random	TRUE	FALSE	FALSE	99% 77%	68%	84%	100
7.	4	hashtag hashtag	random	TRUE	FALSE	TRUE	99%	25%	83%	99
7.	5	hashtag	positive	TRUE	FALSE	TRUE	77%	68%	86%	100
7		hashtag	random	TRUE	TRUE	FALSE	99%	26%	82%	91
7	7 2	hashtag	positive	TRUE	TRUE	FALSE	75%	67%	86%	100
7		hashtag	random	TRUE	TRUE	TRUE	99%	28%	81%	99
7		hashtag	positive	TRUE	TRUE	TRUE	74%	67%	85%	100
8		plain text	random	FALSE	FALSE	FALSE	100%	3%	97%	100
8		plain text	positive	FALSE	FALSE	FALSE	95%	74%	91%	100
8		plain text	random	FALSE	FALSE	TRUE	100%	4%	96%	100
8	3 2	plain text	positive	FALSE	FALSE	TRUE	93%	71%	91%	100
8	4 2	plain text	random	FALSE	TRUE	FALSE	99%	4%	96%	100
8	5 2	plain text	positive	FALSE	TRUE	FALSE	90%	69%	91%	100
8	6 2	plain text	random	FALSE	TRUE	TRUE	99%	10%	89%	95
8	7 2	plain text	positive	FALSE	TRUE	TRUE	83%	65%	90%	100
8	8 2	plain text	random	TRUE	FALSE	FALSE	99%	21%	86%	99
8	9 2	plain text	positive	TRUE	FALSE	FALSE	79%	68%	88%	10
9		plain text	random	TRUE	FALSE	TRUE	99%	23%	85%	99
9	1 2	plain text	positive	TRUE	FALSE	TRUE	78%	67%	88%	100
9.	2 2	plain text	random	TRUE	TRUE	FALSE	99%	23%	84%	99
9	3 2	plain text	positive	TRUE	TRUE	FALSE	79%	67%	87%	100
9.	2	plain text	random	TRUE	TRUE	TRUE	99%	25%	83%	99
				TRUE	TRUE	TRUE	76%	67%	86%	

testing accuracy	positive sentiment accuracy		forum accuracy
0.996498906	0.3078	0.632	0.966223132
0.899092088	0.8512	0.6356	0.98567042
0.994752187	0.3676	0.5776	0.970317298
0.825174825 0.991540516	0.8206	0.6158	0.988741044
0.991540516	0.384	0.5814	0.977482088
0.737935657	0.7376	0.581	0.984646878
0.977622714	0.4092	0.5672	0.977482088
0.626297578	0.4876	0.5872	0.984646878
0.959158719	0.4632	0.5532	0.977482088
0.655647383		0.603	0.983623337
0.952696078	0.4746	0.5538	0.976458547
0.624074074		0.6026	0.982599795
0.943789887	0.4718	0.5482	0.977482088
0.943789887	0.4718	0.5896	0.982599795
0.601489758	0.4124	0.5896	0.982599795
0.944949619	0.4/52	0.5864	0.977482088
0.615528531	0.4016	0.5864	0.981576254
	0.0968	0.879	0.997952917
0.933974587	0.7866	0.8212	0.996929376
0.995014955	0.1504	0.8036	0.987717503 0.997952917
0.861725939	0.756	0.8288	
0.992918104	0.1822	0.7812	0.986693961
0.784204228	0.7078	0.8322	
0.986833585	0.2438	0.7428	0.986693961
0.683404776	0.6278	0.7662	0.988741044
0.985408117	0.391	0.6946	0.986693961
0.702741703	0.6318	0.733	0.988741044
0.983629237	0.4076	0.6804	0.983623337
0.685115931	0.6144	0.7262	0.988741044
0.98818624	0.4154	0.67	0.984646878
0.67350038	0.613	0.7202	0.98567042
0.98534202	0.4258	0.6584	0.981576254
0.659476117	0.4238	0.7164	0.98567042
0.659476117 0.997463155	0.051	0.9422	0.98567042 0.997952917
0.948141045	0.051	0.9422	0.997952917
0.948141045	0.7564	0.8812	0.998976459
0.990114062	0.0636	0.9268	0.998976459
		0.8782 0.869	
0.994509571 0.830168494	0.1122 0.684	0.869 0.8804	0.990788127
0.991186051	0.18	0.803	0.986693961
0.726002704	0.6362	0.8636	
0.990577889	0.317	0.7652	0.986693961
0.741554054	0.6828	0.7888	0.989764585
0.991216795	0.3262	0.7474	0.984646878
0.722422062	0.6712	0.7788	0.988741044
0.990086207	0.342	0.7456	0.986693961
0.71402439	0.6716	0.7638	0.987717503
0.989771491	0.3598	0.7308	0.986693961
0.73354232	0.678	0.751	0.986693961
0.73354232 0.997409087	0.0408	0.9556	1
0.950682282	0.759	0.8928	0.998976459
0.995712755	0.051	0.9474	0.997952917
0.917398731	0.7266	0.8952	0.998976459
0.992236222	0.0904	0.0002	0.989764585
0.862672432	0.6882	0.902	
0.991965066	0.0002	0.858	0.989764585
0.765504587	0.6314	0.8868	
0.991088725		0.8268	
0.991088725 0.762719704	0.2526 0.6754	0.8268 0.8608	0.988741044 0.997952917
0.990796319	0.2694	0.8074	0.988741044
0.737737738	0.6694	0.8522	0.997952917
0.990499293		0.7948	0.988741044
0.746276323 0.992823457	0.6726	0.8102	0.988741044
	0.3078	0.778	0.986693961
0.731486414 0.99761923	0.6618	0.7954 0.9626	0.987717503
	0.0352		1
0.95308567	0.7444	0.9054	0.998976459
0.996064207	0.0392	0.9572	1
0.926591685	0.7268	0.8976	0.998976459
0.994051024	0.053	0.9468	0.998976459
0.885385821	0.693	0.9056	0.998976459
0.990310396	0.1242	0.8742	0.990788127
0.790547103	0.6322	0.8922	0.998976459
0.990404134	0.2368	0.841	0.988741044
0.773046714	0.681	0.87	
0.989659134	0.2472	0.827	0.988741044
0.772727273	0.678	0.8556	
0.001962042	0.578	0.8330	0.000741044
0.991862042 0.751966682	0.2552	0.823	
0.7519000082	0.6672	0.8376	0.997952917
0.991533766	0.2776	0.8086	0.988741044 0.997952917
0.997138632	0.6694	0.8478	0.99/95291/
0.99/138632		0.9728 0.9112	1
0.949295474	0.7406		1
0.996275853	0.0358	0.963	1
0.929467475	0.7104	0.9106	0.998976459
0.993795769	0.0424	0.9596	1
0.903688525	0.6948	0.9144	0.998976459
0.992125984	0.1038	0.8924	0.990788127
0.830743618	0.6454	0.9014	0.998976459
0.994289669	0.2114	0.861	0.988741044
0.789492623	0.6784	0.8818	
0.992753623	0.2262	0.8482	0.989764585
0.784126984 0.992470156	0.6746	0.8766	
	0.2348	0.8448	0.988741044
0.786503067 0.99420669	0.6742	0.8694	
	0.251	0.8318	0.988741044
0.76265419	0.6712	0.8646	0.997952917
	0.0711	0.0040	