Multi label classification performance (one-vs-one, voting):

Roc-SVM is the only method that supports continuous features!
Bayesian Sets can support them, but not a very trivial extension from our current state!

Very high variability in quality of augmentation between experiments!
Furthermore, no clear evidence of more data benefiting augmentation!
Huge variability in the number of new positives of each class added between different experiments!

Prover data set:

	Precision	Recall	F-Score
Pre-augmentation:			
	0.2088	0.3252	**0.2543**
Post augmentation(Roc-SVM with c=1, smaller c has worse performance):			
	0.2261	0.3406	**0.2718**
	0.2223	0.3376	**0.2681**
	0.1959	0.2522	0.2205
	0.225	0.342	**0.2715**
	0.3013	0.1099	0.161
Using more unlabeled data:			
	0.2257	0.3065	**0.26**
	0.1943	0.3132	0.2398
	0.2324	0.3711	**0.2858**
	0.2051	0.3109	0.2472
	0.195	0.2269	0.2098

Prover data set #2 (test and train data swapped):

Pre augmentation:			
	0.2187	0.3213	**0.2603**
Post augmentation(RocSVM with c = 0.1):			
	0.2419	0.339	**0.2823**
	0.2203	0.2306	0.2253
	0.222	0.3142	0.2602
	0.2147	0.324	0.2583
	0.2439	0.3286	**0.28**
Using more unlabeled data:			
	0.1295	0.0936	**0.1086**
	0.2441	0.3034	**0.2705**
	0.1924	0.3238	0.2413
	0.2246	0.2952	0.2551
	0.2077	0.3311	0.2553