

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

Prover data set:

	Precision	Recall	F-Score
<i>Pre-augmentation:</i>			
	0.2088	0.3252	0.2543
<i>Post augmentation(Roc-SVM with $c=1$, smaller c has worse performance):</i>			
	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

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

<i>Pre augmentation:</i>			
	0.2187	0.3213	0.2603
<i>Post augmentation(RocSVM with $c = 0.1$):</i>			
	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