1. Loaded data using Pandas.read\_data()
2. In Preprocessing
   1. Throw away punctuations.
   2. Use SnowBall stemmer with stopwords: English.
   3. TODO: consider pickle dumping.
   4. TF-IDF Vectorization
3. Resolved issue with multiclass F1. Used f1-score(average=’weighted’)
4. Issue with precision\_recall\_curve() not supporting multiclass.
5. Used only ‘document’ and ‘label’ columns for now.
6. Initial run o/p

Processing 2203 samples with 5 attributes  
Evaluating classifiers

Plotting the results

Accuracy of clf: Linear SVC (F1 score=0.721) = 0.7188208616780045

Accuracy of clf: NuSVC (F1 score=0.729) = 0.7301587301587301

Accuracy of clf: Ada Boost (F1 score=0.589) = 0.6009070294784581

1. TODO
   1. Multiclass Pre/Re (label binarizer)
   2. CV