RF no tuning, just Bag of words

('CV prec', 0.67824798459879421)

('CV rec', 0.56759932170542637)

('CV acc', 0.63648198789353694)

confusion matrix test set

[[145 60]

[101 130]]

('test prec', 0.68421052631578949)

('test rec', 0.56277056277056281)

('test acc', 0.63073394495412849)

('Test spec', 0.0)

('test f1', 0.61757719714964365)

Log reg w time c=0.01

('CV prec', 0.68384436299932649)

('CV rec', 0.55738250968992253)

('CV acc', 0.63643771149413209)

confusion matrix test set

[[147 58]

[ 92 139]]

('test prec', 0.70558375634517767)

('test rec', 0.60173160173160178)

('test acc', 0.65596330275229353)

('Test spec', 0.0)

('test f1', 0.64953271028037396)

RF w time

input max depth parameter 100

input n\_estimators parameter 45

input max\_features parameter 31

input min\_samples\_split parameter 80

C:\Users\kny02\Anaconda2\lib\site-packages\sklearn\metrics\classification.py:1135: UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels with no predicted samples.

'precision', 'predicted', average, warn\_for)

input min\_samples\_leaf parameter 5

random forest

('CV prec', 0.63283590190502392)

('CV rec', 0.81262718023255809)

('CV acc', 0.65748222042426774)

confusion matrix test set

[[ 86 119]

[ 28 203]]

('test prec', 0.63043478260869568)

('test rec', 0.87878787878787878)

('test acc', 0.66284403669724767)

('Test spec', 0.0)

('test f1', 0.73417721518987344)

RF w time, attachment

RFparameters={"max\_depth":100, "n\_estimators":45, "max\_features":31, "min\_samples\_split":80, "min\_samples\_leaf":5}

('CV prec', 0.64056087522933403)

('CV rec', 0.81342054263565888)

('CV acc', 0.66558598275025471)

('CV f1', 0.71671531467756111)

confusion matrix test set

[[117 88]

[ 52 179]]

('test prec', 0.67041198501872656)

('test rec', 0.77489177489177485)

('test acc', 0.67889908256880738)

('Test spec', 0.0)

('test f1', 0.71887550200803207)

W counts:

random forest

('CV prec', 0.51862354814288636)

('CV rec', 1.0)

('CV acc', 0.51862354814288636)

('CV f1', 0.68301791945358314)

confusion matrix test set

[[ 0 205]

[ 0 231]]

('test prec', 0.52981651376146788)

('test rec', 1.0)

('test acc', 0.52981651376146788)

('Test spec', 0.0)

('test f1', 0.6926536731634183)

W counts is worse….

TODO mess w counts and email length. We are probably overprioritizing recall

W email length

input max depth parameter 15

input n\_estimators parameter 16

input max\_features parameter 50

input min\_samples\_split parameter 245

C:\Users\kny02\Anaconda2\lib\site-packages\sklearn\metrics\classification.py:1135: UndefinedMetricWarning: F-score is ill-defined and being set to 0.0 in labels with no predicted samples.

'precision', 'predicted', average, warn\_for)

input min\_samples\_leaf parameter 35

random forest

('CV prec', 0.53308567661996376)

('CV rec', 0.96173086240310079)

('CV acc', 0.54251186012904917)

('CV f1', 0.68595032785163434)

confusion matrix test set

[[ 8 197]

[ 1 230]]

('test prec', 0.53864168618266983)

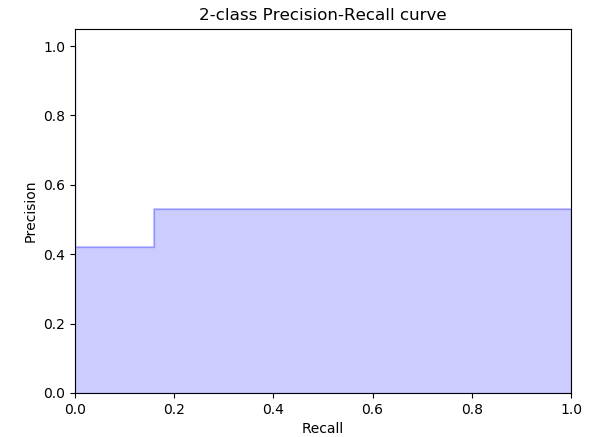
('test rec', 0.99567099567099571)

('test acc', 0.54587155963302747)

('Test spec', 0.0)

('test f1', 0.69908814589665658)

None



Sigmoid SVM

('CV prec', 0.65026000234373571)

('CV rec', 0.71268774224806197)

('CV acc', 0.65182248614637117)

('CV f1', 0.67973379032190828)

confusion matrix test set

[[109 96]

[ 61 170]]

('test prec', 0.63909774436090228)

('test rec', 0.73593073593073588)

('test acc', 0.63990825688073394)

('Test spec', 0.0)

('test f1', 0.68410462776659964)