# Boostexter versus Naïve Bayesian

An implementation of Shapire and Freund’s Boostexter algorithm as applied with AT&T’s “How May I Help You” system versus a Naïve Bayesian Text Classifier.

# Python files

## boostexter.py

The Boostexter algorithm

## boostexter.c

The Boostexter algorithm implemented in C

## naivebayes.py

A Naive Bayesian Text Classifier.

## readutils.py

Utilities to read models and text and create tuples

## cntnorm.py

Procedure to normalize the Boostexter Weights

## classifier.py

Run a clssifier model agian some (labelled or unlabelled) text input

## score.py

Score the model using the Matthews Correlation Coefficient Boostexter only

## scorepos.py

Score the model using the Matthews Correlation Coefficient Boostexter and Naïve Bayesian

## matthews.py

Matthews Correlation Coefficient utilities

## scorecnt.py

Score TP / (TP + FP)

## makeplot.py

Read classifier scores and plot them

## test3.dat

Labeled Wikipedia Pages

## test3\_4.model

Boostexter Model – initial tuples – 5th cross-validation

## test3\_4.model.cntnorm

Normalized Boostexter Model – initial tuples – 5th cross-validation

## test3\_4.model.naivebayes

Naïve Bayesian Model – initial tuples – 5th cross-validation

## scorecnttest\_test3.png

Boostexter, Normalized Boostexter, Naïve Bayesian

Score TP / (TP + FP) versus number of tuples (x 100)

## scorepostest\_test3.png

Boostexter, Normalized Boostexter, Naïve Bayesian

Matthews Correlation Coefficient versus number of tuples (x 100)

## text\_math.dat

Unlabeled math textbook