```
None
Usage: classifier.py [options]

Options:
  -h, --help            show this help message and exit
  --report              Print a detailed classification report.
  --confusion_matrix    Print the confusion matrix.
  --top10               Print ten most discriminative terms per class for
                        every classifier.
  --all_categories      Whether to use all categories or not.
  --use_hashing         Use a hashing vectorizer.
  --n_features=N_FEATURES
                        n_features when using the hashing vectorizer.
  --filtered            Remove newsgroup information that is easily overfit:
                        headers, signatures, and quoting.
Loading 20 newsgroups dataset for categories:
['alt.atheism', 'talk.religion.misc', 'comp.graphics', 'sci.space']
data loaded
2034 documents – 3.979536 MB (training set)
1353 documents – 2.86749 MB (test set)
['alt.atheism', 'talk.religion.misc', 'comp.graphics', 'sci.space']
Extracting features from the training data using a sparse vectorizer
Done in 0.5345475673675537 s at 7.444680778546467 MB/s
No. of Samples : 2034, No. of Features: 33809
Extracting features from the test data using the same vectorizer
Done in 0.32591915130615234s  at 8.798163558380224 MB/s
No. of Samples: 1353, No. of Features: 33809
======================================================================
=========
Perceptron
________________________________________________________________________________
Training:
Perceptron(alpha=0.0001, class_weight=None, eta0=1.0, fit_intercept=True,
      max_iter=None, n_iter=50, n_jobs=1, penalty=None, random_state=0,
      shuffle=True, tol=None, verbose=0, warm_start=False)
/usr/local/lib/python3.5/dist-packages/sklearn/linear_model/stochastic_gradient.py:117:
DeprecationWarning: n_iter parameter is deprecated in 0.19 and will be removed in 0.21. Use
max_iter and tol instead.
  DeprecationWarning)
Train time: 0.10479879379272461s
Test time:  0.0015597343444824219s
accuracy:   0.885439763488544
top 10 keywords per class:
alt.atheism
 mantis example religion okcforum rice religious thing keith atheists atheism
comp.graphics
 bates albany jr0930 pov video windows package imagine file graphics
sci.space
 observations solar comet planets astro moon sci funding orbit space
talk.religion.misc
 loving told mr stephen 2000 frank fbi buffalo abortion christian
classification report:
```

```
                  precision    recall  f1-score   support

      alt.atheism       0.85      0.81      0.83       319
    comp.graphics       0.91      0.96      0.94       389
        sci.space       0.93      0.94      0.94       394
talk.religion.misc       0.80      0.77      0.79       251

      avg / total       0.88      0.89      0.88      1353

confusion matrix:
[[259   8   9  43]
 [  4 373   9   3]
 [  5  15 372   2]
 [ 36  12   9 194]]
================================================================================
=========
Naive Baye's assuming Multinomial Distribution
________________________________________________________________________________
Training:
MultinomialNB(alpha=0.01, class_prior=None, fit_prior=True)
Train time: 0.007391691207885742s
Test time:  0.0016553401947021484s
accuracy:   0.8994826311899483
top 10 keywords per class:
alt.atheism
 say livesey article don people atheists com caltech god keith
comp.graphics
 file image com files nntp host posting thanks university graphics
sci.space
 pat moon digex henry article access gov com nasa space
talk.religion.misc
 apple don kent article people sandvik jesus christian god com
classification report:
                  precision    recall  f1-score   support

      alt.atheism       0.85      0.87      0.86       319
    comp.graphics       0.95      0.95      0.95       389
        sci.space       0.92      0.95      0.94       394
talk.religion.misc       0.86      0.77      0.81       251

      avg / total       0.90      0.90      0.90      1353

confusion matrix:
[[279   2   8  30]
 [  2 369  16   2]
 [  3  15 376   0]
 [ 45   4   9 193]]
================================================================================
=========
Naive Baye's assuming Bernoulli Distribution
________________________________________________________________________________
Training:
```

```
BernoulliNB(alpha=0.01, binarize=0.0, class_prior=None, fit_prior=True)
Train time: 0.008842945098876953s
Test time:  0.007508516311645508s
accuracy:   0.8839615668883961
top 10 keywords per class:
alt.atheism
 god say think people don com nntp host posting article
comp.graphics
 like com article know thanks graphics university nntp host posting
sci.space
 nasa like university just com nntp host posting space article
talk.religion.misc
 think know christian posting god people just don article com
classification report:
                    precision    recall  f1-score   support

       alt.atheism       0.83      0.88      0.86       319
     comp.graphics       0.88      0.96      0.92       389
         sci.space       0.94      0.91      0.92       394
talk.religion.misc       0.87      0.73      0.79       251

       avg / total       0.88      0.88      0.88      1353

confusion matrix:
[[282   9   3  25]
 [  1 373  13   2]
 [  5  31 358   0]
 [ 50  10   8 183]]
================================================================================
K Nearest Neighbours
________________________________________________________________________________
Training:
KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
           metric_params=None, n_jobs=1, n_neighbors=10, p=2,
           weights='uniform')
Train time: 0.0011320114135742188s
Test time:  0.16659164428710938s
accuracy:   0.8580931263858093
top 10 keywords per class:
Top features cannot be Selected
confusion matrix:
[[287   3  11  18]
 [ 14 348  19   8]
 [  7  26 359   2]
 [ 59  13  12 167]]
```