#
Configuration 1
#
Run with default parameters: a training dataset and a test dataset.
#
By default, the classification method used is a SciKit Pipeline based on a Multinomial NaiveBayes classifier with TF-IDF weighting and feature selection (k=500).
#
Both SVMs and Multinomial NaiveBayes were also tested, with slightly worse results.

```
python mini-project.py --train training.csv --test test.csv
```

```
fold-1: training...
fold-1: classifying...
fold-1: evaluation...
--- START OF EVALUATION ---
        ----- RUN 1 -----
        PRECISION: 0.698717948718
        ACCURACY: 0.663888888889
        RECALL: 0.598901098901
        F-MEASURE: 0.644970414201
        ----- END OF RUN 1 -----

AVERAGE PRECISION: 0.698717948718

--- END OF EVALUATION ---
```

#
Configuration 2
#
Run with a training dataset, a test dataset, and a list of confounding features the system must avoid.

```
python mini-project.py —train training.csv --test test.csv --exceptions exceptions.txt
```

```
fold-1: training...
fold-1: classifying...
fold-1: evaluation...
--- START OF EVALUATION ---
        ----- RUN 1 -----
        PRECISION: 0.830508474576
        ACCURACY: 0.6
        RECALL: 0.269230769231
        F-MEASURE: 0.406639004149
        ----- END OF RUN 1 -----

AVERAGE PRECISION: 0.830508474576

--- END OF EVALUATION ---
```

#
Configuration 3
#
Run with a training dataset, a test dataset, a list of confounding features the system must avoid, and frequency pruning of features (f=10).
#
This configuration yields the lowest recall but also the highest accuracy. F-measure is very negatively affected, though, so the previous configuration is probably the optimal one.

```
python mini-project.py --train training.csv --test test.csv --exceptions exceptions.txt --prune 10
```

```
fold-1: training...
fold-1: classifying...
fold-1: evaluation...
--- START OF EVALUATION ---
        ----- RUN 1 -----
        PRECISION: 0.941176470588
        ACCURACY: 0.533333333333
        RECALL: 0.0879120879121
        F-MEASURE: 0.160804020101
        ----- END OF RUN 1 -----

AVERAGE PRECISION: 0.941176470588

--- END OF EVALUATION ---
```

#

Configuration 4

#

Run with a single dataset for training and testing (10-fold cross-validation will be used to evaluate the results), and the list of confounding features the system must avoid.

#

Average precision in this configuration is slightly worse than using the test dataset, suggesting the test dataset may be slightly biased.

#

Precision also greatly varies across folds, suggesting some level of internal structure to the input collection (when using cross-validation, the dataset is randomized, which means that the observed effects should be occurring dataset-wide). Theres's probably some form of clustering that could be used to improve the data homogeneity and, hence, the classification accuracy.

```
python mini-project.py --train ../data/training.csv --exceptions exceptions.txt
```

```
fold-1: training...
fold-1: classifying...
fold-1: evaluation...
fold-2: training...
fold-2: classifying...
fold-2: evaluation...
fold-3: training...
fold-3: classifying...
fold-3: evaluation...
fold-4: training...
fold-4: classifying...
fold-4: evaluation...
fold-5: training...
fold-5: classifying...
fold-5: evaluation...
fold-6: training...
fold-6: classifying...
fold-6: evaluation...
fold-7: training...
fold-7: classifying...
fold-7: evaluation...
fold-8: training...
fold-8: classifying...
fold-8: evaluation...
fold-9: training...
fold-9: classifying...
fold-9: evaluation...
fold-10: training...
fold-10: classifying...
fold-10: evaluation...
--- START OF EVALUATION ---
        ----- RUN 1 -----
        PRECISION: 0.797101449275
        ACCURACY: 0.615384615385
        RECALL: 0.273631840796
        F-MEASURE: 0.407407407407
        ----- END OF RUN 1 -----
```

----- RUN 2 -----
PRECISION: 0.8
ACCURACY: 0.59375
RECALL: 0.234146341463
F-MEASURE: 0.362264150943
----- END OF RUN 2 -----

----- RUN 3 -----
PRECISION: 0.888888888889
ACCURACY: 0.579326923077
RECALL: 0.221198156682
F-MEASURE: 0.354243542435
----- END OF RUN 3 -----

----- RUN 4 -----
PRECISION: 0.548895899054
ACCURACY: 0.608173076923
RECALL: 0.896907216495
F-MEASURE: 0.681017612524
----- END OF RUN 4 -----

----- RUN 5 -----
PRECISION: 0.904761904762
ACCURACY: 0.649038461538
RECALL: 0.355140186916
F-MEASURE: 0.510067114094
----- END OF RUN 5 -----

----- RUN 6 -----
PRECISION: 0.8375
ACCURACY: 0.581730769231
RECALL: 0.295154185022
F-MEASURE: 0.436482084691
----- END OF RUN 6 ——

----- RUN 7 -----
PRECISION: 0.551724137931
ACCURACY: 0.612980769231
RECALL: 0.907216494845
F-MEASURE: 0.686159844055
----- END OF RUN 7 -----

----- RUN 8 -----
PRECISION: 0.847222222222
ACCURACY: 0.603365384615
RECALL: 0.283720930233
F-MEASURE: 0.425087108014
----- END OF RUN 8 ——

----- RUN 9 -----
PRECISION: 0.561904761905
ACCURACY: 0.625
RECALL: 0.907692307692
F-MEASURE: 0.694117647059
----- END OF RUN 9 -----

```
        ----- RUN 10 -----
        PRECISION: 0.791044776119
        ACCURACY: 0.598557692308
        RECALL: 0.257281553398
        F-MEASURE: 0.388278388278
        ----- END OF RUN 10 -----

AVERAGE PRECISION: 0.752904404016

--- END OF EVALUATION ---
```