# Load and prepare the dataset

import nltk

from nltk.corpus import movie\_reviews

import random

documents = [(list(movie\_reviews.words(fileid)), category)

for category in movie\_reviews.categories()

for fileid in movie\_reviews.fileids(category)]

random.shuffle(documents)

# Define the feature extractor

all\_words = nltk.FreqDist(w.lower() for w in movie\_reviews.words())

word\_features = list(all\_words)[:2000]

def document\_features(document):

document\_words = set(document)

features = {}

for word in word\_features:

features['contains({})'.format(word)] = (word in document\_words)

return features

# Train Naive Bayes classifier

featuresets = [(document\_features(d), c) for (d,c) in documents]

train\_set, test\_set = featuresets[100:], featuresets[:100]

classifier = nltk.NaiveBayesClassifier.train(train\_set)

# Test the classifier

print(nltk.classify.accuracy(classifier, test\_set))

classifier.show\_most\_informative\_features(5)