Laboratorium 4 - klasyfikacja dokumentów tekstowych

W NLTK jest zaimplementowany klasyfikator NaiveBayesClassifier. Sprawdzić jak mozna użyc tego klasyfikatora. Wykonać eksperymenty na zbiorze

1. opinii o filmach, zawartych w korpusie movie_reviews. Dane można znalezć w katalogu corpora, podkatalog movie_reviews, podkatalog pos lub neg, lub
2. NPS Chat Corpus, który zawiera 10 000 postów z etykietami "Statement", "Emotion" itd

Dla potrzeb testów prosze zmieniac następujace parametry:

- liczba słów branych pod uwagę w uczeniu i testowaniu spośród najczęściej występujących słów w korpusie,
- zbiór treningowy i testowy powinien być zbiorem słów pochodzących z pewnego dokumentu z korpusu(zbiory rozłaczne!)
- rozmiar zbioru treningowego i testowego -

Prosze wykonać eksperymenty sprawdzające jak zmieni się dokładność klasyfikacji dla różnych wartości parametrów.
Informacje na temat klasyfikacji mozna znalezć w Natural Language Processing with Python - Steven Bird, Ewan Klein, and Edward Loper rozdz.6.
Opisane wyniki prosze przesłać w pliku **nazwiskoLab4.pdf**.