Homework 3 - Feature Engineering
2015 Spring, Machine Learning
Choong-Wan Woo
Kaggle Username: CU_chwo9116
February 12, 2015

Explanation.

1) I removed any accent symbols using the **strip_accents** option of the CountVectorizer.
2) I removed numbers and punctuations using the **preprocessor** option of the CountVectorizer.
3) Using the **tokenizer** option and WordNetLemmatizer provided by the nltk package, I did a lemmatization.
4) I removed english stopwords using the **stop_words** option of the CountVectorizer.
5) I created an additional feature that consists of two words in a row in the document using the **ngram_range** option of the CountVectorizer.