Day 2

* Looked at Naïve-Bayes classification algorithm
* Used porter’s algorithm for stemming English.

To stem a word

Source: <http://web.eecs.umich.edu/~radev/coursera-slides/nlpintro_co2_02.04_DR_Edit.pdf>

* Did an example using Naïve-Bayes algorithm
* Chapter 2 Important Definitions
  + The word error rate
    - “the normalized minimum edit distance in words between our output and the gold: the number of word insertions, deletions, and substitutions divided by the length of the gold sentence in words;” – The Book
  + Lemmatization
    - Task of determining that two words have the same root, despite their surface differences.
    - Done through a technique called **morphological parsing.**
    - **Morphology:** the study of the way words are built up from smaller meaning-bearing units called **morphemes.**
    - **Stems**
      * Central morpheme of the word
    - **Affixes**
      * Adding “additional” meanings of various kinds.
* Also learned about Normalization in Chapter 2 of *Speech and Language Processing*
* Delved deeper into Named-Entity Recognition (NER) through an NLTK tutorial series.