This is the explanation for Kaggle competition.

The implementation includes the implementation for Random classifier, Naive base

classifier and Naive base classifier with Laplas smoothing

In the implementation there are three phases:

1. Data loader:
   1. Load data
   2. Using NLTK to filter words and only keep alphabetical words
   3. Filter words with less than two characters
   4. Convert all words to lower case
   5. Stemming
2. Naïve base model
   1. Develop the naïve base model
   2. Develop the naïve base model with laplas smoothing
      1. I tried different alpha and the best one was alpha = 2.15
3. Prediction
   1. Predict the outcome of the model