How to make it even better

Play around with tokenization

- Special tokens like emoji, ":)" and "!!!" can help

Try to normalize tokens

- Adding stemming or lemmatization

Try different models

- SVM, Naïve Bayes, …

Throw BOW away and use Deep Learning

- https://arxiv.org/pdf/1512.08183.pdf
- Accuracy on test set in 2016: 92.14% (+2.5%)