# Web Scraping and NLP

- Spam classification with Naive Bayes
- Twitter API for NLP feature construction (ngrams, etc) to classify Democrat/Republican politicians + sentiment analysis
- Scrape all machine learning pages on Wikipedia https://en.wikipedia.org/wiki/Category:Machine_learn and run latent Dirichlet allocation

## Spam classification with Naive Bayes

## Analysis of US politicians' Tweets

## Latent Dirichlet allocation on Wikipedia pages