Naive Bayes for Text Classification

Jupyter notebooks

- Vectorization
- A SMS spam detector
- Classifying messages by topic

Homework (due, Oct 12, Monday)

- Yelp reviews
- Revisiting the linguistic street map of Singapore
- Email spam detection

In-class project (due, Oct 12, Monday)

- The Disputed Federalist Papers

Datasets

Filename	Description	Source
sms	Collection of SMS messages tagged as spam or ham (legitimate)	Kaggle.com
yelp.csv	Collection of Yelp reviews	Kaggle.com
emails	Collection of emails tagged as spam or ham	Apache SpamAssassin's public datasets
messages.csv	Collection of emails tagged by topic	scikit-learn
papers.txt	Federalist Papers	Gutenberg project