

Detecting “Fake” News with NLP

Teresa Borcuch

Goal: to identify defining linguistic and textual characteristics of fake news, reputable news, and satire

Stakeholders:

- journalists
- policy makers
- social media
- advertisers
- consumers of media



Data Sources

The Real:

newsapi.org

The Fake:

.co sites

Satire:



THE BOROWITZ REPORT
Not the news.

Bloomberg

bloomberg



BUSINESS
INSIDER

business-insider



Analysis

Features to Consider:

- Word frequency in article/headlines
- Content matching of headline vs. text
- Structure of headline (? or not)
- Average word/sentence length
- Parts of Speech
- Others?

* restricting task to language processing only
(disregard meta-information like comments sections,
posting details, etc)

Comments, Questions, Concerns?