NATURAL LANGUAGE PROCESSING OF ECONOMIC NEWS ARTICLES

Steven Bierer

Metis Data Science, Seattle

DATA

- 8000 brief articles and headlines (1951-2014)
  - source: figure-eight.com/data-for-everyone/
- Respondents noted relevance to U.S. economy
- Respondents judged tone (negative=1, positive=10)

APPROACH

- SpaCy to tokenize words
  - named entities: organization + location
  - categories: monetary and percentage values
- Gensim for latent dirichlet allocation
  - 15 topics (10,000 word dictionary)
- Separate modeling of articles and headlines
- Sklearn for classification and regression

RESULTS - LDA TOPIC EXAMPLES

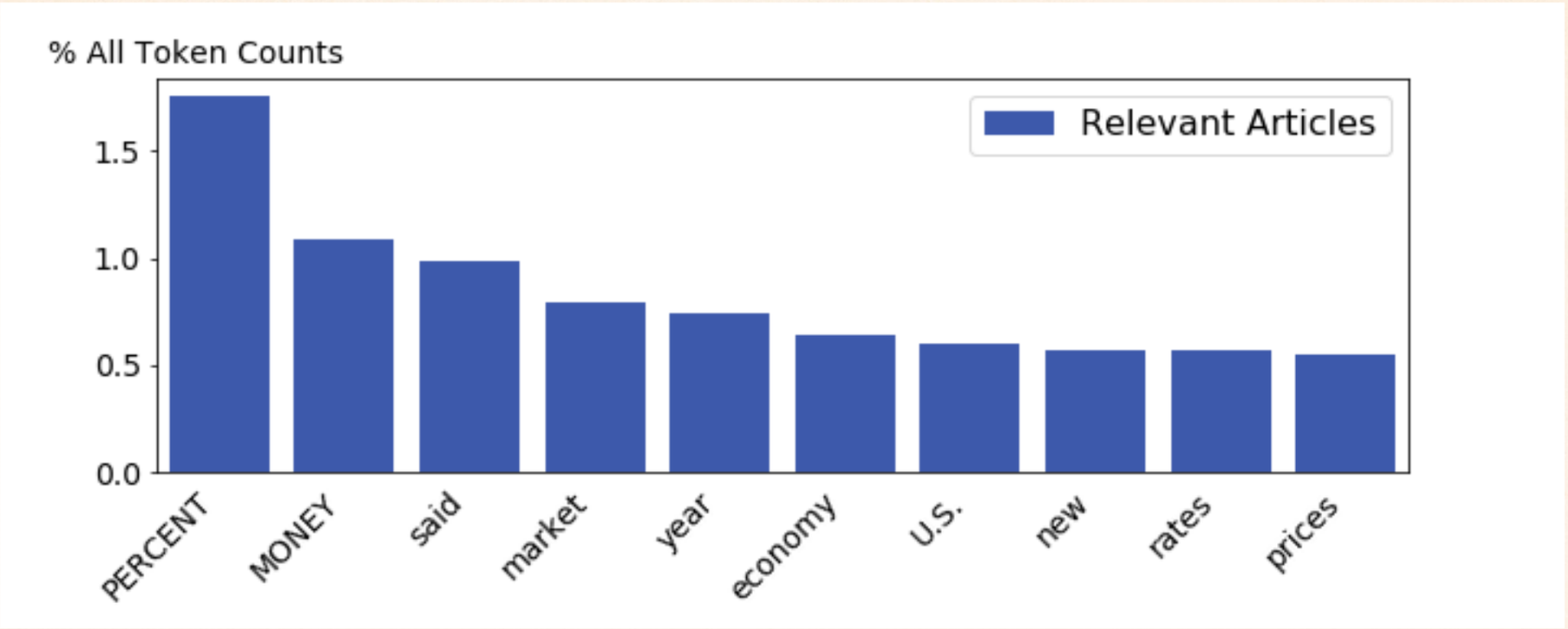
```
'0.035*"budget" +
0.031*"tax" +
0.031*"deficit" +
0.022*"MONEY" +
0.020*"cut" +
0.016*"spending" +
0.016*"government" +
0.014*"federal" +
0.012*"year" +
0.011*"fiscal" +
0.009*"president" +
0.008*"administration" +
0.008*"say" +
0.008*"CONGRESS" +
0.008*"congress"'
```

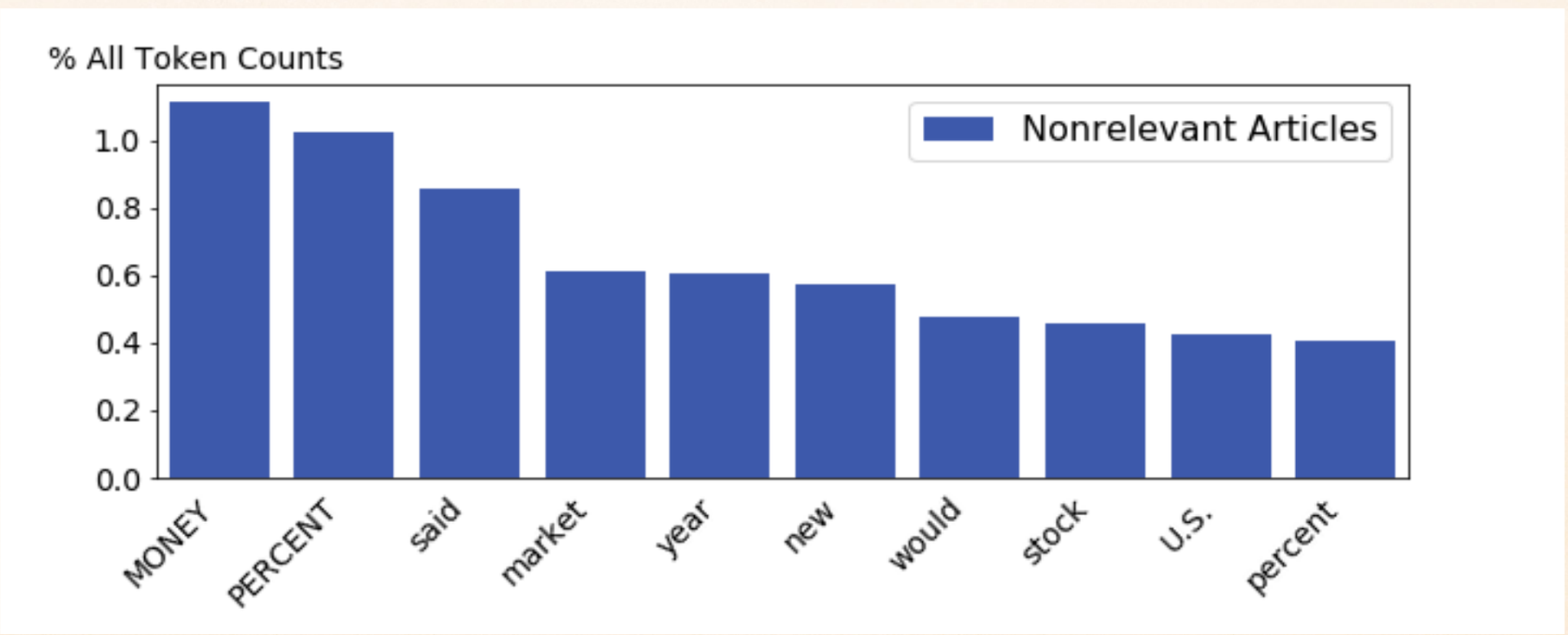
```
'0.016*"president" +
0.014*"say" +
0.012*"bush" +
0.008*"republican" +
0.008*"political" +
0.007*"economic" +
0.007*"campaign" +
0.006*"make" +
0.006*"state" +
0.006*"election" +
0.006*"administration" +
0.005*"party" +
0.005*"democrat" +
0.005*"democratic" +
0.005*"get"'
```

RESULTS - PREDICTIVE MODELS

- Article classification (relevant/nonrelevant)
  - KNN or Random Forest
  - Accuracy = .81; precision = .30; recall = .070
  - Terrible! (18% relevant articles)

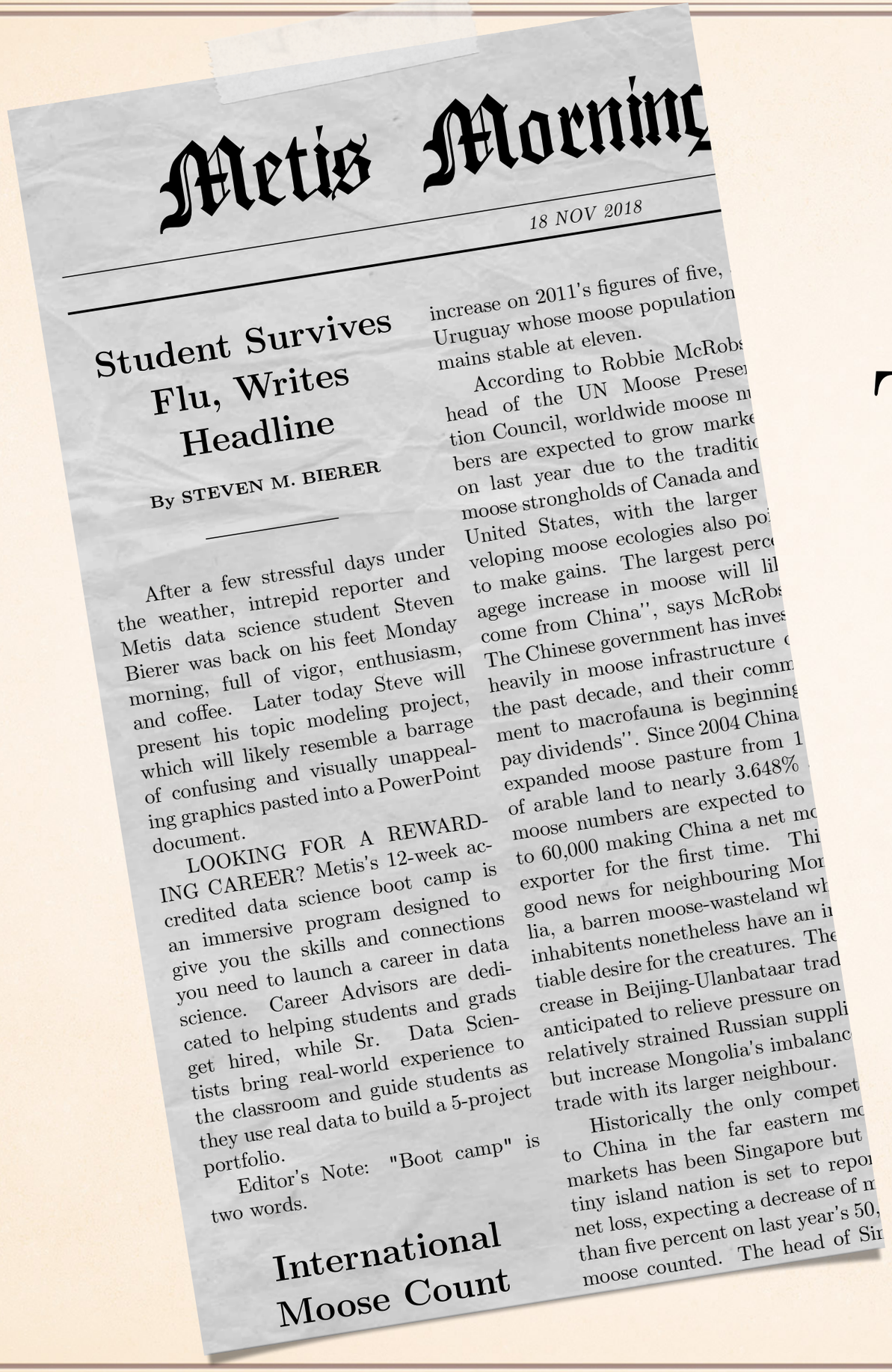
- Article positivity
  - Linear Regression
  - R-squared = 0.04
  - Terrible!


% All Token Counts
Relevant Articles
1.5
1.0
0.5
0.0
PERCENT
MONEY
said
market
year
economy
U.S.
new
rates
prices


% All Token Counts
Nonrelevant Articles
1.0
0.8
0.6
0.4
0.2
0.0
MONEY
PERCENT
said
market
year
new
would
stock
U.S.
percent

EDITOR'S NOTE

- Remove most common words on economy
- Exploit named entities in U.S.
  - States, companies, institutions, stock symbols
- Apply sentiment analysis (easy w/ SpaCy)



Metis Morning

18 NOV 2018

Student Survives Flu, Writes Headline

By STEVEN M. BIERER

After a few stressful days under the weather, intrepid reporter and Metis data science student Steven Bierer was back on his feet Monday morning, full of vigor, enthusiasm, and coffee. Later today Steve will present his topic modeling project, which will likely resemble a barrage of confusing and visually unappealing graphics pasted into a PowerPoint document.

LOOKING FOR A REWARDING CAREER? Metis's 12-week accredited data science boot camp is an immersive program designed to give you the skills and connections you need to launch a career in data science. Career Advisors are dedicated to helping students and grads get hired, while Sr. Data Scientists bring real-world experience to the classroom and guide students as they use real data to build a 5-project portfolio.

Editor's Note: "Boot camp" is two words.

International Moose Count

increase on 2011's figures of five, Uruguay whose moose population mains stable at eleven.

According to Robbie McRobs head of the UN Moose Prese tion Council, worldwide moose n bers are expected to grow marke on last year due to the traditio moose strongholds of Canada and United States, with the larger veloping moose ecologies also po to make gains. The largest perc agege increase in moose will lil come from China", says McRobs The Chinese government has inves heavily in moose infrastructure the past decade, and their comm ment to macrofauna is beginning pay dividends". Since 2004 China expanded moose pasture from 1 of arable land to nearly 3.648% moose numbers are expected to to 60,000 making China a net mo exporter for the first time. Thi good news for neighbouring Mor lia, a barren moose-wasteland wh inhabitents nonetheless have an i tiable desire for the creatures. The crease in Beijing-Ulanbataar trad anticipated to relieve pressure on relatively strained Russian suppli but increase Mongolia's imbalanc trade with its larger neighbour.

Historically the only compet to China in the far eastern mo markets has been Singapore but tiny island nation is set to repo net loss, expecting a decrease of n than five percent on last year's 50, moose counted. The head of Sin

THANK YOU