

A-NEWS-MENT (FORMERLY SPOT-A-BOT) PROPOSAL

Sentiment Analysis Project

By Andrew Yaspan, Juan Arroyo Miranda, Ratul Esrar

INTRODUCTION

“In a world of alternative truths, even the language used in unbiased news reports is seldom neutral.” – Joe Biden

- Better understanding of sentiment in news articles
- A-News-Ment
 - Web app
 - Choose a topic from dynamic menu of trending topics
 - Search keywords
 - Scrape USA Today and La Jornada for related/relevant articles
 - Generate sentiment scores for text from said articles
 - Goal is to be accurate on sentiment $\geq 60\%$ of time

PROJECT DESCRIPTION & GOALS

Develop a sentiment measure that allows us to classify reporter or quote sentiment in NEWSPAPER ARTICLES

- Scraping USA TODAY (tentatively) AND LA JORNADA (mexican newspaper)
- We will use positive or negative sentiment tag lists to determine sentiment content

Deliverable

- A web application that will take user input, parse the input for keywords, and then return a sentiment analysis for articles about applicable keywords.

Possible Extension

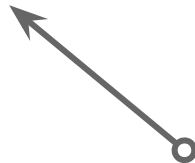
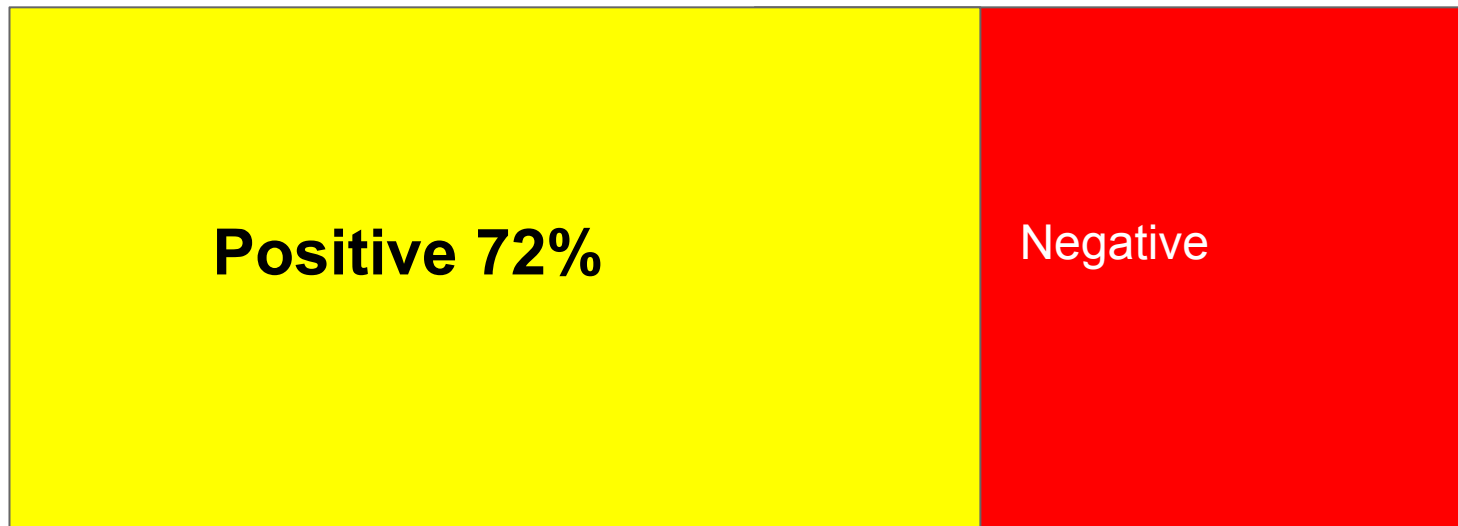
- Correlate news sentiment to historical financial indicators

SCRAPING EXAMPLE

```
beautifulLab.py task7.sql Indices.csv prices.csv companies.csv create-comp.sql
37 html = pm.urlopen(url=my_url, method="GET").data
38 soup = bs4.BeautifulSoup(html)
39
40
41 student@CS-Vbox: ~/cs122-win-17-ayaspan
42 ion actresses of all time with two timeless hits: <a href="/topic/entertainment/television/the-dick-van-dyke-s
43 how-%28tv-program%29-ENTTV000000295-topic.html">&quot;The Dick Van Dyke Show&quot;</a> and &quot;The Mary Tyle
44 r Moore Show.&quot; Since then, she's become known for her guest roles, as well as for her extensive charity w
45 ork.Moore was born in 1936 in Brooklyn Heights, but her family moved to Los Angeles when she was 8 years old.
46 As a teenager, she aspired to be a dancer and appeared in several commercials at the beginning of her career.<
47 /p></div> <div class="trb embed related credit and caption"><p><a href="/topic/entertainment/mary-tyler-moo
48 re-PECLB003254-topic.html">Mary Tyler Moore</a> cemented her status as one of the great television actresses o
49 f all time with two timeless hits: <a href="/topic/entertainment/television/the-dick-van-dyke-show-%28tv-progr
50 am%29-ENTTV000000295-topic.html">&quot;The Dick Van Dyke Show&quot;</a> and &quot;The Mary Tyler Moore Show.&
51 uot; Since then, she's become known for her guest roles, as well as for her extensive charity work.Moore was b
52 orn in 1936 in Brooklyn Heights, but her family moved to Los Angeles when she was 8 years old. As a teenager,
53 she aspired to be a dancer and appeared in several commercials at the beginning of her career.</p> (Carolyn Co
54 le / Los Angeles Times)</div> <div class="trb_sc" data-role="sc_information_sc_container
55 " data-sc-c="page" data-sc-url="/entertainment/tv/lat-moore6_lyf5yipd20120126144647-photo.html" data-sc-ti="Ma
56 ry Tyler Moore: Career in pictures" data-sc-th="http://www.trbimg.com/img-5720d67b/turbine/lat-moore6_lyf5yipd
57 20120126144647-thumbnail" data-sc-desc="Mary Tyler Moore cemented her status as one of the great television ac
58 tresses of all time with two timeless hits: &quot;The Dick Van Dyke Show&quot; and &quot;The Mary Tyler Moore
59 Show.&quot; Since then, she&apos; become known for her guest roles, as well as for her extensive charity work
60 .Moore was born in 1936 in Brooklyn..." data-sc-sl="lat-moore6_lyf5yipd20120126144647" data-sc-cont="photoGall
61 ery" data-sc-nn="Los Angeles Times" data-sc-contid="67661991"><span class="trb_sc_item" data-role="sc_share" d
62 ata-sc-t="facebook" ></span> <span class="trb_sc_item" data-role="sc_share" data-sc-t="twitter" ></span> <sp
63 an class="trb_sc_item" data-role="sc_email" data-sc-t="email" ></span> </div> <a class="trb_lightbox_mor
64 eGalleriesLink" href="/visuals/photography/">See more galleries</a></div> <div class="trb_gptAd" dat
65 a-role="adloader adloader_ad" data-adloader-size="300x250" data-state="adloader_unloaded" data-withinviewport
66 -options="bottomOffset=100" data-adloader-position="lightBoxCube" data-adloader-adtype="cube" data-adloader-p
```

Line 61, Column 1 Spaces: 4 Python

SENTIMENT SCORING MOCK-UP



ANTICIPATED TIMELINE

1. **Data Collection** (week 5)
 - a. Scrape news websites
 - b. News API to gather additional news
 - c. Yahoo API to gather information on financial indicators
2. **Data Cleaning** (week 7)
3. **Sentiment index development** (week 8)
 - a. Scrape text of newspaper sites and count presence of words in the positive and negative word lists
 - b. Construct positive and negative word lists for Spanish
4. **User Interface** (week 9)
5. Tentative correlations with financial indicators