



News Article Sentiment Analysis

What's the process

Part 1
Find 2 news
articles

- Look for 2 or more opposing views
- Find topics to compare

Part 2
Process top
stories/RSS
feed

- Find new articles and topics to analyze – at least 5 days

Part 3
Web
scrape
articles

- Store data in a remote DB
- Extract data to one row for analysis

Part 4
Sentimental
analysis and
visualization

- Store sentiment query results in DB
- Use 3 sentiment points for visualization

Data Sources

AllSides™ Media Bias Chart™

All ratings are based on online content only — not TV, print, or radio content. Ratings do not reflect accuracy or credibility; they reflect perspective only.



ALTERNET
BuzzFeed NEWS
CNN
DEMOCRACY NOW!
DAILY BEAST
HUFFPOST
The Intercept
JACOBIN
Mother Jones
msnbc
THE NEW YORKER
The New York opinion Times
NATION.
SLATE
Vox

abc
AP politics & fact check
The Atlantic
Bloomberg
CBS
The Economist
theguardian
NBC
The New York news Times
npr opinion
POLITICO
PROPUBLICA
TIME
The Washington Post
USA TODAY
YAHOO! NEWS

AP
AXIOS
BBC
THE CHRISTIAN SCIENCE MONITOR
IJR. INDEPENDENT JOURNAL REVIEW
MarketWatch
Newsweek
npr news
REUTERS
REAL CLEAR POLITICS
THE HILL
THE WALL STREET JOURNAL news

The American Conservative
DeseretNews.
THE DISPATCH
THE EPOCH TIMES
newsmax news
NEW YORK POST news
The Post Millennial.
reason
THE WALL STREET opinion JOURNAL
Examiner
The Washington Times

THE AMERICAN SPECTATOR
BREitbart
BLAZE
CBN
THE DAILY CALLER
Daily Mail
DAILY WIRE
FOX NEWS
FEDERALIST
NATIONAL REVIEW
NEW YORK POST opinion
newsmax opinion
OAN

LEFT LEAN LEFT CENTER LEAN RIGHT RIGHT

AllSides Media Bias Ratings™ are based on multi-partisan, scientific analysis. Visit [AllSides.com](https://www.AllSides.com) to view hundreds of media bias ratings.

Version 5 | AllSides 2021

Technologies & Libraries used



Technologies

- Python
- Flask
- Bootstrap
- MongoDB Atlas
- Jupyter Notebook
- Power BI

Libraries

- BeautifulSoup
- PyMongo
- Flask
- pandas

1.

News Agencies & Articles

CNN & Fox News



LEFT WING



RIGHT WING

2.

Process News Articles

CNN & Fox News



www.cnn.com



www.foxnews.com

3.

Web Scrape Articles

CNN

```
In [1]: from bs4 import BeautifulSoup
import pandas as pd
import re
import feedparser
import string

def clean_text(raw_html):
    cleantext = BeautifulSoup(raw_html, "html.parser").text
    cleantext = cleantext.replace('\n', '')
    return cleantext
```

```
rss_url = 'http://rss.cnn.com/rss/cnn_topstories.rss'
cols = ['title', 'summary', 'link']

news_df = pd.DataFrame(columns = cols)
feed = feedparser.parse(rss_url)
feed
```

```
feed.entries[0].summary
'72WNTAKBA' /></a> <a href="http://rss.cnn.com/~ff/rss/cnn_topstories?i=6usaLMBak4A:D4q1z69oLI:V_sGLIPBpMU"><img border="0" s
rc="http://feeds.feedburner.com/~ff/rss/cnn_topstories?i=6usaLMBak4A:D4q1z69oLI:V_sGLIPBpMU" /></a> <a href="http://rss.cnn.
com/~ff/rss/cnn_topstories?i=6usaLMBak4A:D4q1z69oLI:qj6IDK7rITs"></a> <a href="http://rss.cnn.com/~ff/rss/cnn_topstories?i=6usaLMBak4A:D4q1z69oLI:gIN9vFwOqVQ">
</a></div>
>,
'links': [{'rel': 'alternate',
'type': 'text/html',
'href': 'http://rss.cnn.com/~r/rss/cnn_topstories/-3/6usaLMBak4A/index.html'}],
'link': 'http://rss.cnn.com/~r/rss/cnn_topstories/-3/6usaLMBak4A/index.html',
'id': 'https://www.cnn.com/2021/11/02/politics/virginia-governor-race-nationalization/index.html',
'guidislink': False,
'published': '2021-11-02T13:40:05-05:00'}
```

CNN Feed Scraping



```
3 Three states you should be watching on electio... Elections in Virginia, New York and Minnesota ... https://www.cnn.com/videos/politics/2021/10/29...
4 Opinion: The answer Virginia voters deserve fr... Virginia's Republican candidate for governor, ... https://www.cnn.com/2021/10/31/opinions/trump...
```

```
In [3]: # Sentiment score from vader
from nltk.sentiment import SentimentIntensityAnalyzer
def get_vader_sentiment(txt):
    analyzer = SentimentIntensityAnalyzer()
    vs = analyzer.polarity_scores(txt)
    return vs['compound']

news_df['compound'] = news_df['summary'].apply(lambda x: (get_vader_sentiment(x)))
news_df.head()
```

	title	summary	link	compound
0	If Youngkin wins, he could serve as a template...	Analysis: Why Democrats are in a lot of trou...	https://www.cnn.com/2021/11/02/politics/virgin...	-0.4019
1	Polls open for races that will test the poll...	What to watch in today's elections Analysis...	https://www.cnn.com/politics/live-news/electo...	0.5859
2	What to watch in today's elections	Tuesday's elections in states and cities acros...	https://www.cnn.com/2021/11/02/politics/us-eli...	0.0000
3	Three states you should be watching on electio...	Elections in Virginia, New York and Minnesota ...	https://www.cnn.com/videos/politics/2021/10/29...	0.0000
4	Opinion: The answer Virginia voters deserve fr...	Virginia's Republican candidate for governor, ...	https://www.cnn.com/2021/10/31/opinions/trump...	0.7964

```
In [4]: news_df['sentiment'] = news_df['compound'].apply(lambda c: 'neutral' if c==0 else 'positive' if c>0 else 'negative')
news_df.head()
```

	title	summary	link	compound	sentiment
0	Opinion: The answer Virginia voters deserve fr...	Virginia's Republican candidate for governor, ...	https://www.cnn.com/2021/10/31/opinions/trump...	0.7964	positive
3	Three states you should be watching on electio...	Elections in Virginia, New York and Minnesota ...	https://www.cnn.com/videos/politics/2021/10/29...	0.0000	neutral
5	What to watch in today's elections	Tuesday's elections in states and cities acros...	https://www.cnn.com/2021/11/02/politics/us-eli...	0.0000	neutral
4	Opinion: The answer Virginia voters deserve fr...	Virginia's Republican candidate for governor, ...	https://www.cnn.com/2021/10/31/opinions/trump...	0.7964	positive
0	If Youngkin wins, he could serve as a template...	Analysis: Why Democrats are in a lot of trou...	https://www.cnn.com/2021/11/02/politics/virgin...	-0.4019	negative

```
Out[4]:
news_df.head()

In [4]: news_df['sentiment'] = news_df['compound'].apply(lambda c: 'neutral' if c==0 else 'positive' if c>0 else 'negative')
news_df.head()
```

CNN Sentiment Analysis

CNN

```
In [5]: import pymongo
        from pymongo import MongoClient

        cluster = MongoClient("mongodb+srv://testuser:testuser123@cluster0.dj0nr.mongodb.net/myFirstDatabase?retryWrites=true&w=majority")
        db = cluster["sentimentAnalysisDB"]
        collection = db["cnnnewsData"]

        collection.insert_many(news_df.to_dict('records'))

Out[5]: <pymongo.results.InsertManyResult at 0x21714624d00>

In [6]: cursor = db["cnnnewsData"].find()

        # Expand the cursor and construct the DataFrame
        df = pd.DataFrame(list(cursor))
        df.to_csv('cnn.csv', index=False)
```

CNN Data Transfer to & from MongoDB



Fox News

```
In [6]: from bs4 import BeautifulSoup
import pandas as pd
import re
import feedparser
import string

def clean_text(raw_html):
    cleantext = BeautifulSoup(raw_html, "html.parser").text
    # cleantext = BeautifulSoup(raw_html, "lxml").text
    return cleantext

rss_url = 'http://feeds.foxnews.com/foxnews/latest'
cols = ['title', 'summary', 'published', 'link']

news_df = pd.DataFrame(columns = cols)
feed = feedparser.parse(rss_url)

for newssitem in feed['items']:
    news_df = news_df.append({'title': newssitem['title'],
                             'summary': clean_text(newssitem['summary']),
                             'published': newssitem['published'],
                             'link': newssitem['link']], ignore_index=True)

news_df.head()
```

```
Out[6]:
```

	title	summary	published	link
0	Prior COVID-19 infection does not protect as w...	Vaccine-induced immunity provides more protect...	Wed, 03 Nov 2021 03:10:55 GMT	https://www.foxnews.com/health/prior-covid-19-...
1	MSNBC's Joy Reid: 'Education' issue is 'code' ...	MSNBC's Joy Reid called the issue of education...	Wed, 03 Nov 2021 03:02:18 GMT	https://www.foxnews.com/media/joy-reid-educati...
2	Rev. Jesse Jackson discharged from hospital af...	The Rev. Jesse Jackson was discharged from the...	Wed, 03 Nov 2021 02:59:59 GMT	https://www.foxnews.com/politics/rev-jesse-jac...

```
In [7]: # Sentiment score from Vader
from nltk.sentiment import SentimentIntensityAnalyzer
def get_vader_sentiment(txt):
    analyzer = SentimentIntensityAnalyzer()
    vs = analyzer.polarity_scores(txt)
    return vs['compound']

news_df['compound'] = news_df['summary'].apply(lambda x : (get_vader_sentiment(x)))
news_df.head()
```

```
Out[7]:
```

	title	summary	published	link	compound
0	Prior COVID-19 infection does not protect as w...	Vaccine-induced immunity provides more protect...	Wed, 03 Nov 2021 03:10:55 GMT	https://www.foxnews.com/health/prior-covid-19-...	0.0000
1	MSNBC's Joy Reid: 'Education' issue is 'code' ...	MSNBC's Joy Reid called the issue of education...	Wed, 03 Nov 2021 03:02:18 GMT	https://www.foxnews.com/media/joy-reid-educati...	0.5541
2	Rev. Jesse Jackson discharged from hospital af...	The Rev. Jesse Jackson was discharged from the...	Wed, 03 Nov 2021 02:59:59 GMT	https://www.foxnews.com/politics/rev-jesse-jac...	-0.1531
3	VA Dept. of Elections commissioner says electi...	Virginia Department of Elections Commissioner ...	Wed, 03 Nov 2021 02:53:28 GMT	https://www.foxnews.com/politics/va-dept-of-el...	0.0000
4	Winsome Sears, Virginia GOP lieutenant govern...	Winsome Sears, the Republican candidate for Vi...	Wed, 03 Nov 2021 02:49:47 GMT	https://www.foxnews.com/politics/winsome-sears...	0.1280

```
In [8]: news_df['sentiment'] = news_df['compound'].apply(lambda c: 'neutral' if c==0 else 'positive' if c>=0.01 else 'negative')
news_df.head()
```

```
Out[8]:
```

	title	summary	published	link	compound	sentiment
0	Prior COVID-19 infection does not protect as w...	Vaccine-induced immunity provides more protect...	Wed, 03 Nov 2021 03:10:55 GMT	https://www.foxnews.com/health/prior-covid-19-...	0.0000	neutral
1	MSNBC's Joy Reid: 'Education' issue is 'code' ...	MSNBC's Joy Reid called the issue of education...	Wed, 03 Nov 2021 03:02:18 GMT	https://www.foxnews.com/media/joy-reid-educati...	0.5541	positive
2	Rev. Jesse Jackson discharged from hospital af...	The Rev. Jesse Jackson was discharged from the...	Wed, 03 Nov 2021 02:59:59 GMT	https://www.foxnews.com/politics/rev-jesse-jac...	-0.1531	negative

Fox News Feed Scraping

Fox News Sentiment Analysis



Fox News

```
In [4]: import pymongo
        from pymongo import MongoClient

        cluster = MongoClient("mongodb+srv://testuser:testuser123@cluster0.dj0nr.mongodb.net/myFirstDatabase?retryWrites=true&w=majority")
        db = cluster["sentimentAnalysisDB"]
        collection = db["foxnewsData"]

        collection.insert_many(news_df.to_dict('records'))
```

```
Out[4]: <pymongo.results.InsertManyResult at 0x14fe377aa40>
```

```
In [5]: cursor = db["foxnewsData"].find()

        # Expand the cursor and construct the DataFrame
        df = pd.DataFrame(list(cursor))
        df.to_csv('fox.csv', index=False)
```

Fox News Data Transfer to & from MongoDB



MongoDB

This screenshot shows the MongoDB Atlas interface for a project named 'Project 1'. The left sidebar contains navigation options: DEPLOYMENT, Databases, Triggers, Data Lake, SECURITY, Database Access, Network Access, and Advanced. The 'Databases' section is expanded, showing a tree view with 'sentimentAnalysisDB' and its sub-database 'cnnewsData'. The main panel displays the 'sentimentAnalysisDB.cnnewsData' collection, with a collection size of 82.83KB, 207 total documents, and 36KB of indexes. A filter bar at the top shows a query: `{ field: 'value' }`. Below the filter, the 'QUERY RESULTS 1-20 OF MANY' are displayed as JSON documents. The first document is:

```
{
  "_id": "ObjectId('6170558e7cfaf7af7dab81')",
  "title": "A solar flare could create a geomagnetic storm and put on a stunning s...",
  "summary": "A large solar flare erupted Thursday and is set to reach Earth Saturday...",
  "link": "https://www.cnn.com/2021/10/25/weather/northern-southern-lights-us-eur...",
  "compound": 0.5267,
  "sentiment": "positive"
}
```

The second document is:

```
{
  "_id": "ObjectId('6170558e7cfaf7af7dab82')",
  "title": "DC-Baltimore region braces for tidal flooding that could be the worst ...",
  "summary": "A powerful storm is flooding some Mid-Atlantic coastal areas Friday ev...",
  "link": "https://www.cnn.com/2021/10/25/weather/dc-region-coastal-flooding/index...",
  "compound": -0.3182,
  "sentiment": "negative"
}
```

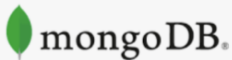
This screenshot shows the MongoDB Atlas interface for the same project, but displaying the 'sentimentAnalysisDB.foxnewsData' collection. The collection size is 14.32KB, with 50 total documents and 36KB of indexes. The filter bar shows the same query: `{ field: 'value' }`. The 'QUERY RESULTS 1-20 OF MANY' are displayed as JSON documents. The first document is:

```
{
  "_id": "ObjectId('6170558e7cfaf7af7dab83')",
  "title": "FBI Baltimore warns: Northern Virginia police increase security at sh...",
  "summary": "Police in Northern Virginia are ramping up security in public spaces s...",
  "published": "Sat, 30 Oct 2021 02:11:51 GMT",
  "link": "https://www.foxnews.com/baltimore-northern-virginia-security-malls-ba...",
  "compound": 0.1827,
  "sentiment": "positive"
}
```

The second document is:

```
{
  "_id": "ObjectId('6170558e7cfaf7af7dab84')",
  "title": "Amy Hurman says 'The View' has an 'unbearable culture' that rewards...",
  "summary": "Former 'View' co-host Amy Hurman is speaking out about what led to...",
  "published": "Sat, 30 Oct 2021 02:03:49 GMT",
  "link": "https://www.foxnews.com/bellaamy-hurman-the-view-unbearable-cult...",
  "compound": 0,
  "sentiment": "neutral"
}
```

MongoDB Data Records



4.

Challenges in Sentimental Analysis

Challenges of Sentiment Analysis and its Accuracy

- Irony and sarcasm
- Types of negations
- Word ambiguity
- Multipolarity



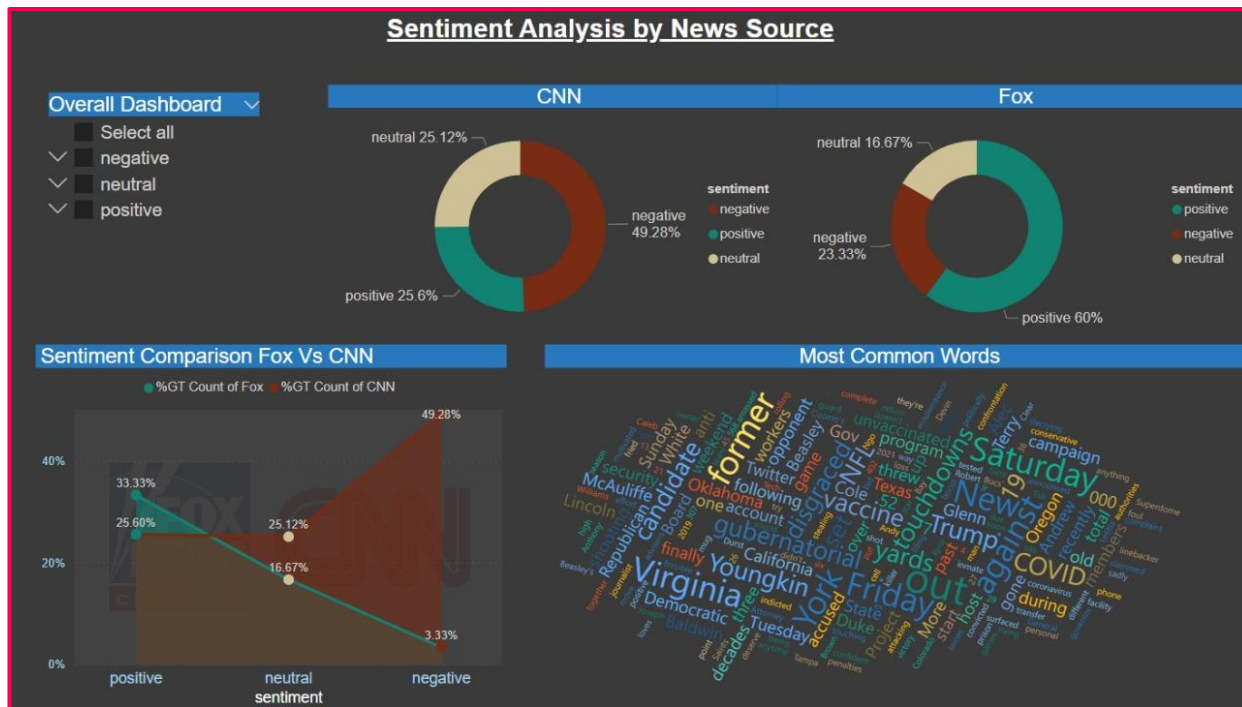
Addressing the challenges of Sentiment Analysis and its Accuracy

Strong sentiment analysis API.

Wide range of AI-based techniques such as text analytics, natural language processing for analysis.

Improved machine learning - Evaluating the training data for bias and updating the labels.

Final Analysis Visualization using Power BI



Conclusion

With this sentiment analysis we can conclude that CNN articles are inclined towards negative where Fox news articles are more positive.

Based on visualization, it seems that the articles that are extremely negative in CNN news while in Fox articles are very positive.

Overall, based on the sentiment comparison graph we could able to cross reference both CNN & Fox, and able to see that CNN gave almost half of CNN articles are Negative whereas half of Fox articles are either positive or Neutral.

From the world cloud it can be identified that the recent governor election in Virginia has been the hot topic and the elected governor Youngkin have been on the discussion.

Fox has been taking side to conservative and CNN towards Liberal can also been able to verified. The publication on the novel COVID has been reported either neutral or positive from both channels.

Thank You...