

Real-Time News Sentiment Analysis Pipeline

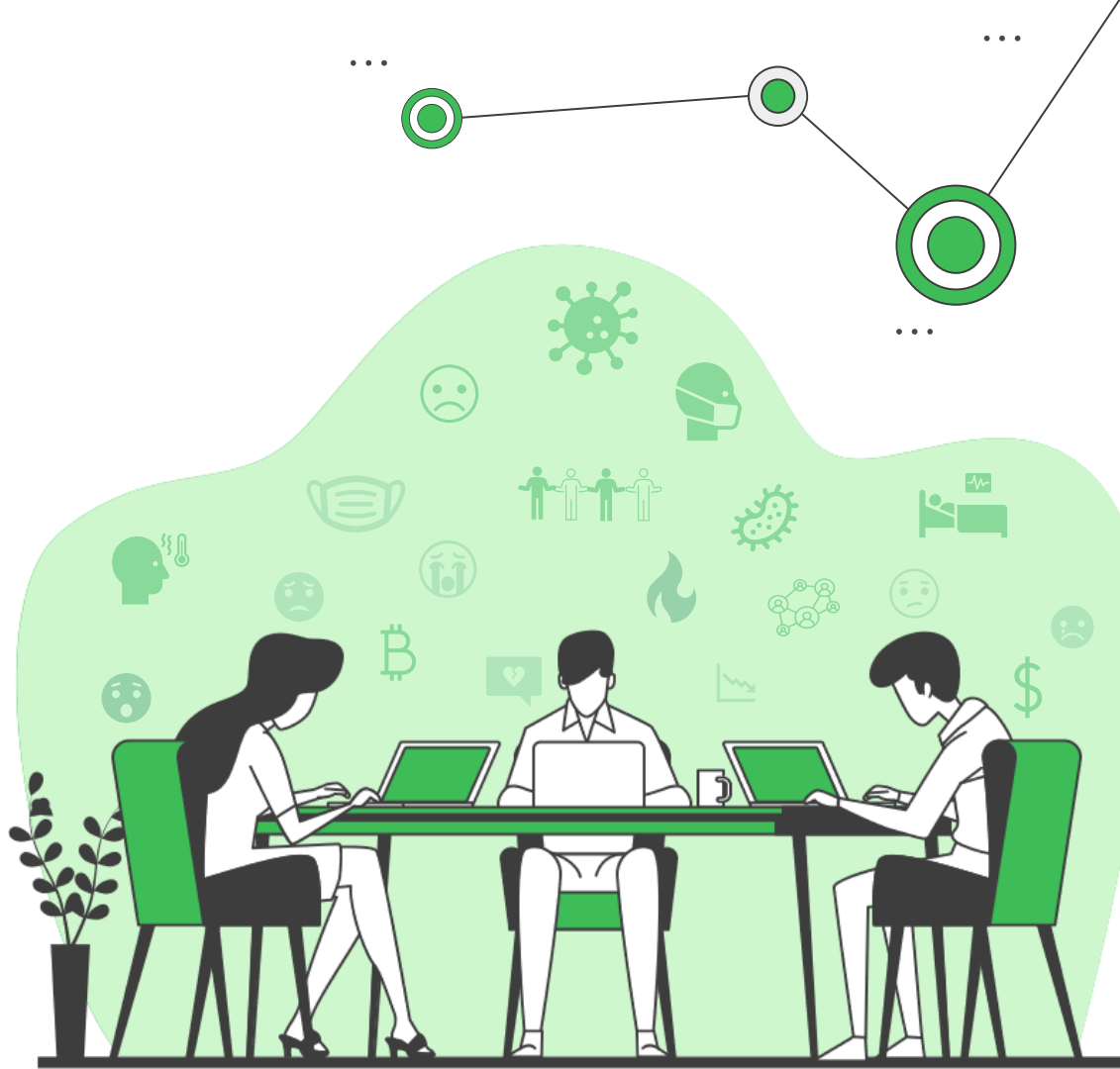
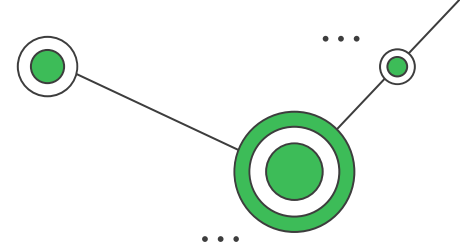
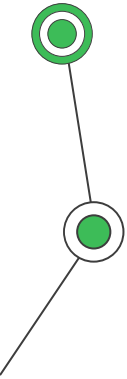


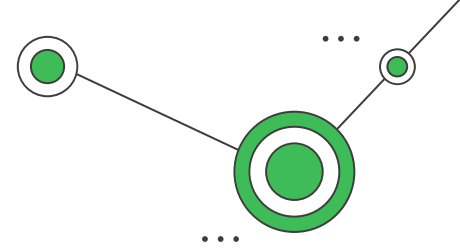
Table of Contents



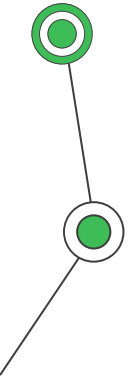
- Objective
- Tools
- What Is Sentiment Analysis?
- Data
- Data Infrastructure
- Sentiment Results
- Real-Time Sentiment Analysis Dashboard
- Conclusion & Next Steps



Objective



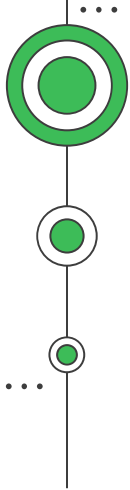
With all the negativity in news headlines recently, it can be a real strain on mental health when you are constantly bombarded with negative news articles and headlines. This project was intended to assist, when you just need a break from negative news articles and want to filter news by positive, neutral or negative sentiment. This was the problem I wanted to create a solution for and by developing a pipeline that performs real-time sentiment analysis on news articles and allows for the user to filter results based on positive or negative sentiment of each article.



Tools

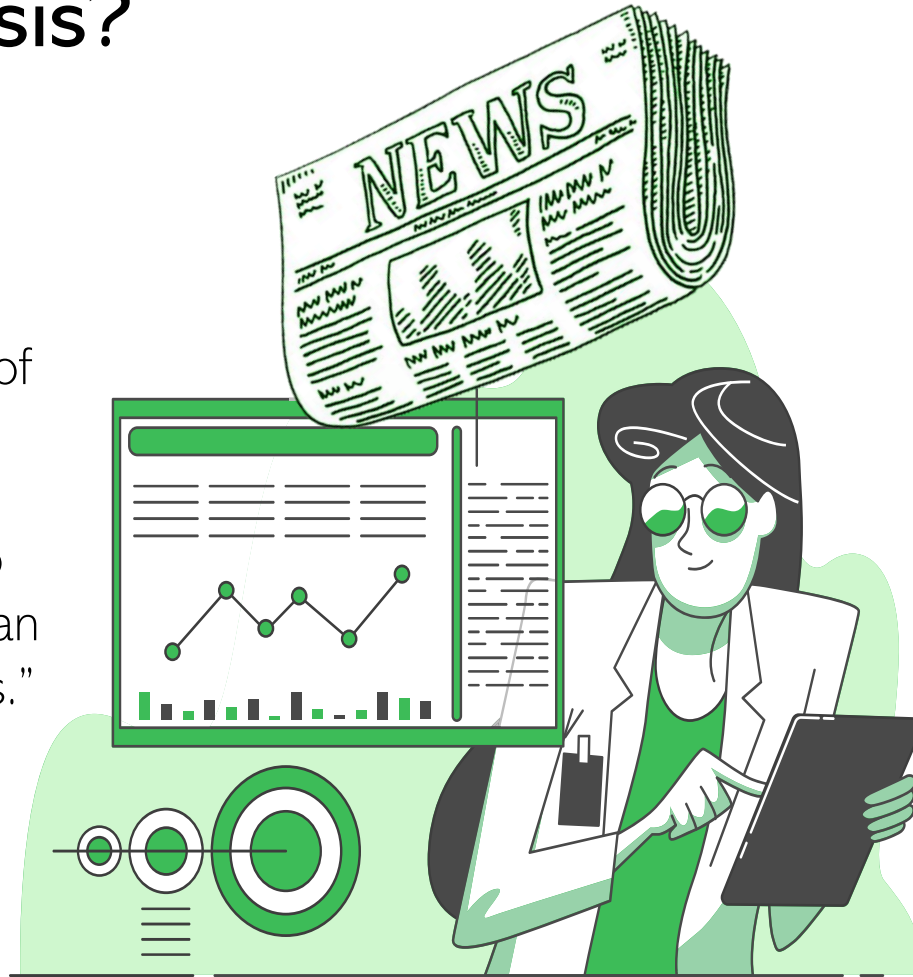


...



What is Sentiment Analysis?

“Sentiment Analysis is a computational process to determine the overall sentiment of a piece of text. This sentiment is usually categorized into 3 main output’s, Positive, Neutral and Negative. This process can help determine the “mood” of the text and this can be extremely useful in a variety of use cases.”



Data

The New York Times

The screenshot shows the New York Times homepage for Tuesday, January 19, 2021. The main headline is "McConnell, Weighing Impeachment, Says Trump 'Provoked' Capitol Mob". Other articles include "Listen to 'The Daily'", "Listen to 'Sway'", and "A Look Back at Trump's Presidency in 6 Minutes". A video player is also visible with the title "A Look Back at Trump's Presidency in 6 Minutes".

*Data is returned as a Json object

{T} The New York Times Developers

{T} Developers

Times Newswire API

Times Newswire API

With the Times Newswire API, you can get links and metadata for Times' articles as soon as they are published on NYTimes.com. The Times Newswire API provides an up-to-the-minute stream of published articles. You can filter results by source (all, nyt, inyt) and section (arts, business, ...).

```
/content/{source}/{section}.json
```

Example Calls

```
https://api.nytimes.com/svc/news/v3/content/all/all.json?api-key=yourkey
https://api.nytimes.com/svc/news/v3/content/ny/business.json?api-key=yourkey
https://api.nytimes.com/svc/news/v3/content/ny/world.json?api-key=yourkey
```

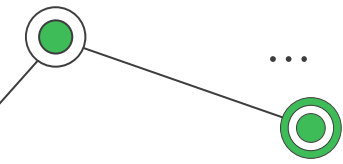
The section-list endpoint returns a list of sections.

```
/content/section-list
```

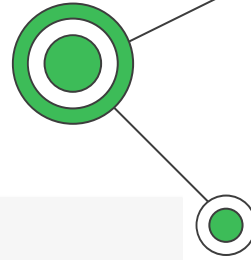
Example Call

```
https://api.nytimes.com/svc/news/v3/content/section-list.json?api-key=yourkey
```

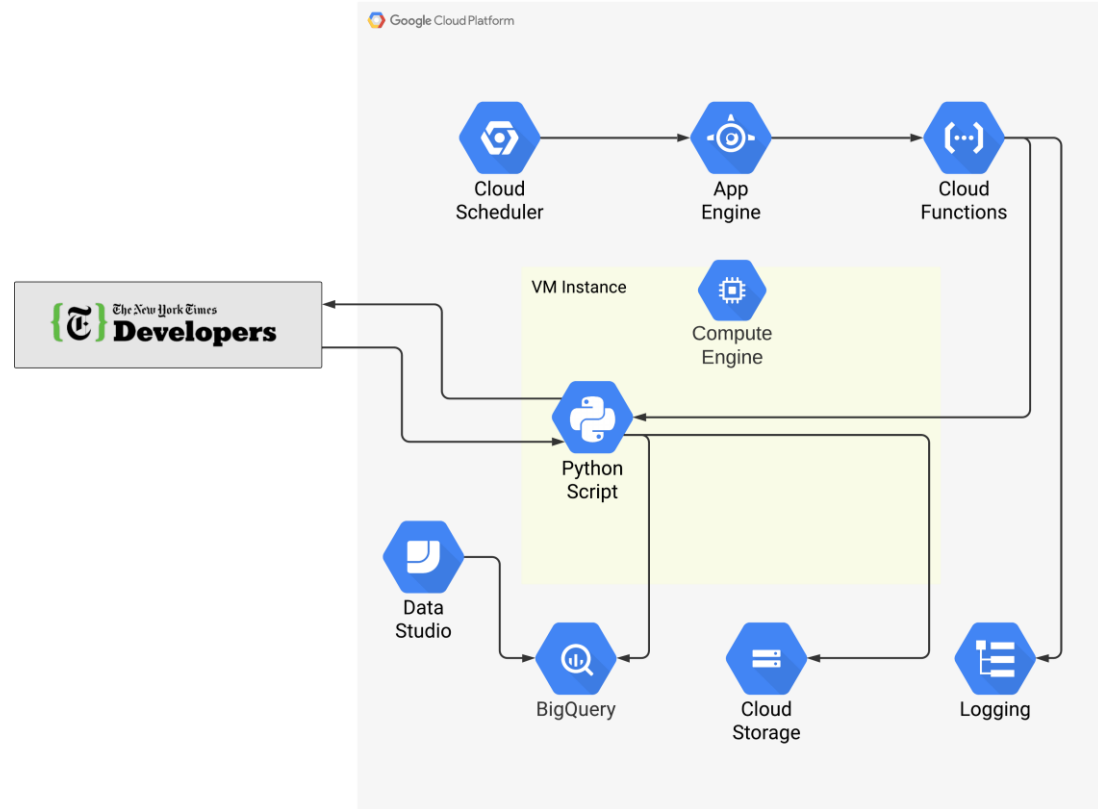
Tiger Woods Announces He Had Fifth Back Operation	Sports	19tiger	BY ELENA BERGERON	Article	News	['Golf']	['PGA Tour Inc']	['Woods, Tiger']	null	Woods, 45, has had five operations on his spine in the past seven years. He expects to miss the first two months of the PGA Tour.	2021-01-19 22:15:24 UTC	https://www.nytimes.com/2021/01/19/sports/golf/tiger-woods-back-surgery.html
---	--------	---------	-------------------	---------	------	----------	------------------	------------------	------	---	-------------------------	--



Data Infrastructure



Cloud Scheduler is used to schedule the pipeline execution via a cronjob and using HTTP App Engine to run a Cloud Function which will execute a python script hosted on a VM Compute Engine and will log the function run data. The script returns news headlines data from the real-time Timeswire API. Next the script processes the data and utilizes VADER Sentiment analysis to analyze the title and abstract columns. Creates a aggregated sentiment score for each article. Then the data is stored as a CSV file and uploaded to a GCP Cloud Storage Bucket and also the data is loaded into a BigQuery table. Finally the BigQuery table is accessed by the DataStudio dashboard and data is refreshed every few minutes, for a real-time sentiment analysis dashboard.



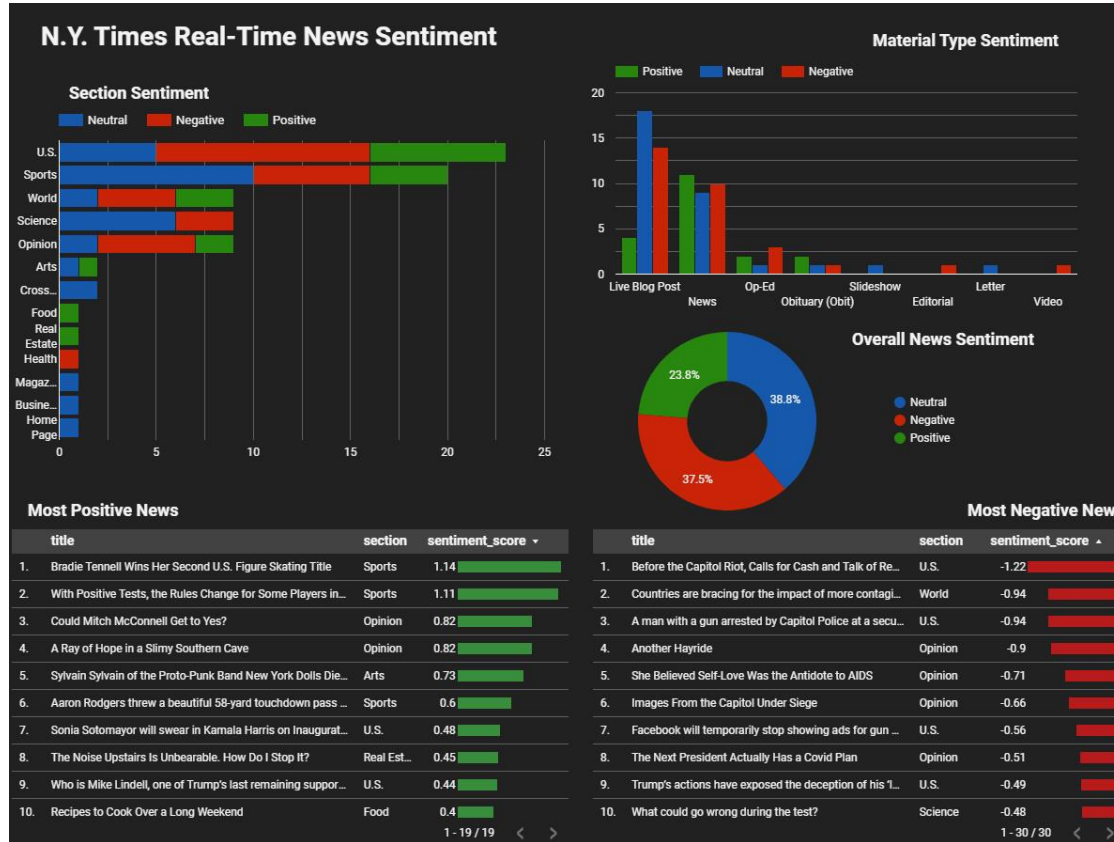
Sentiment Results

	title	section	slug_name	byline	item_type	material_type_facet	des_facet	org_facet	per_facet	geo_facet
0	Deepening Schism, McConnell Says Trump 'Provok...	U.S.	19dc-impeach	BY NICHOLAS FANDOS	Article	News	[United States Politics and Government, Stormi...	[Senate, Republican Party]	[McConnell, Mitch, Trump, Donald J]	None
1	NASA's Decision for Moon Rocket: Test Again or...	Science	19nasa	BY KENNETH CHANG	Article	News	[Rocket Science and Propulsion, United States ...	[National Aeronautics and Space Administration]	[Bridenstine, James F (1975-)]	None
2	Biden's Peloton bike could present cybersecuri...	U.S.	19transition-briefing-biden-peloton	BY SHERYL GAY STOLBERG	Article	Live Blog Post	None	None	None	None
3	Biden to Announce Broad Plan to Reverse Trump ...	U.S.	19dc-immig	BY MICHAEL D. SHEAR	Article	News	[Illegal Immigration, Immigration and Emigrati...	None	[Biden, Joseph R Jr, Trump, Donald J]	None
4	Parler Tries to Survive With Help From Russian...	Technology	19parler	BY JACK NICAS	Article	News	[Right-Wing Extremism and Alt-Right, Computers...	[Amazon.com Inc, Parler LLC]	[Matze, John, Trump, Donald J]	None

compound_title	neg_title	neu_title	pos_title	compound_ab	neg_ab	neu_ab	pos_ab	sentiment_score	sentiment
	0.0000	0.000	1.000	0.000	0.2023	0.062	0.850	0.088	0.2023 Positive
	0.0000	0.000	1.000	0.000	0.2023	0.095	0.773	0.132	0.2023 Positive
	-0.2732	0.174	0.826	0.000	0.0000	0.000	0.000	0.000	-0.2732 Negative
	0.0000	0.000	1.000	0.000	0.6705	0.000	0.800	0.200	0.6705 Positive
	0.4019	0.000	0.748	0.252	0.6369	0.065	0.690	0.246	1.0388 Positive

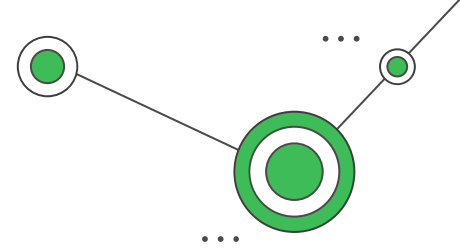
VADER Sentiment Analysis Results

Real-Time Sentiment Dashboard



*Click Here
For Live
Dashboard*

Conclusion & Next Steps



Conclusion:

In conclusion the project was successful and I was able to make use of the real-time news sentiment analysis and although the model does work fairly well, I am looking forward to improving the sentiment analysis results.

Next Steps:

- The next steps for this project would be to create a Streamlit app to be able to access the article's and for a more user friendly application.
- Improving the Sentiment model using Google Natural Language Processing API

