

Web Social Media Analytics and Visualizations
Report

Oyelakin O. Emmanuel
(20115037)

Date

10th May, 2020

Course title

**CMP7202 Web Social Media Analytics and
Visualizations**

Topic modelling and sentiment analysis on social media data: Tweets analysis

Problem statement

- a. What are popular trends on Twitter at the moment, either in the UK or worldwide? Extract some insights from these trends.
- b. Apply sentiment analysis to show users' opinions about one of the selected trends on Twitter. Represent your findings using statistical descriptive methods.

Background

The recent coronavirus (COVID-19) has created a global emergency demanding social distancing, working from home, and self-quarantining to control the spread of the virus. More people have turned to social media to share their thoughts and communicate about this pandemic. It's important to uncover topics being shared on social media platforms relating to COVID-19. Analyzing such information can help policymakers and health care organizations assess the needs of people.

Data

For this project, I used twitter's search application programming interface (API), "Tweepy" Python library. I used this library to obtain 29000 English tweets around the globe related to "coronavirus" search keyword between March 29th and May 9th in 2020. I extracted text and metadata (user profile name, location, mobile device type and tweet timestamp). Figure 1 shows a snippet of the data.

	text	user	source	location	created_at
0	The Mexican government reporting hundred , pos...	nytimes	SocialFlow	New York City	2020-05-08 12:00:13
1	Coronavirus called " Democratic hoax " @ realD...	AprilDRyan	iPhone	Washington, D.C.	2020-05-09 14:46:08
2	Roy Horn Siegfried Roy Has Died 75 After Catch...	TMZ	SocialFlow	Los Angeles, CA	2020-05-09 01:26:33
3	[9th May] (Worldwide) Coronavirus disease ...	DataChronology	Web App		2020-05-10 02:04:08
4	RT @ briantylcohen : And . 72 Wisconsinites ...	toughgal66	Android	New York, USA	2020-05-10 02:04:08
5	RT @ ASlavitt : 5x higher Denmark , 11x higher...	InnovationPith	Android	USA	2020-05-10 02:04:06
6	RT @ minhntgo : Some call karma , I call scien...	shasimss	iPad	Hell	2020-05-10 02:04:06
7	RT @ dwnews : Experts warned suspending mass v...	cefestellita1	Web App		2020-05-10 02:04:05
8	RT @ jamesperloff : Bill Gates explaining need...	blackbird9_9	Web App	North Carolina, USA	2020-05-10 02:04:05
9	RT @ denise4607 : Elon Musk sue county COVID-1...	trfindley	Android	California, USA	2020-05-10 02:04:04

Figure 1-Twitter Data Snippet

Preprocessing Data

The data went through a series of pre-processing steps. I removed punctuation, stop words and nonprintable characters such as emojis from the tweets.

Furthermore, various forms of the same word (e.g., run, running) were lemmatized by converting them to their root word using the “WordNetLemmatizer” module of the Natural Language Toolkit Python library.

Data Analysis

The processed tweets were analyzed and the tweet sources were plotted to show what device users use in tweeting. We can see in Figure 2 that users tweet more with an android device (32%) followed closely by an iPhone (30%) then the web app (23%). This shows that about more than half of users on twitter access the site on their mobile device, twitter should place priority on making sure user experience on their mobile app is very good.

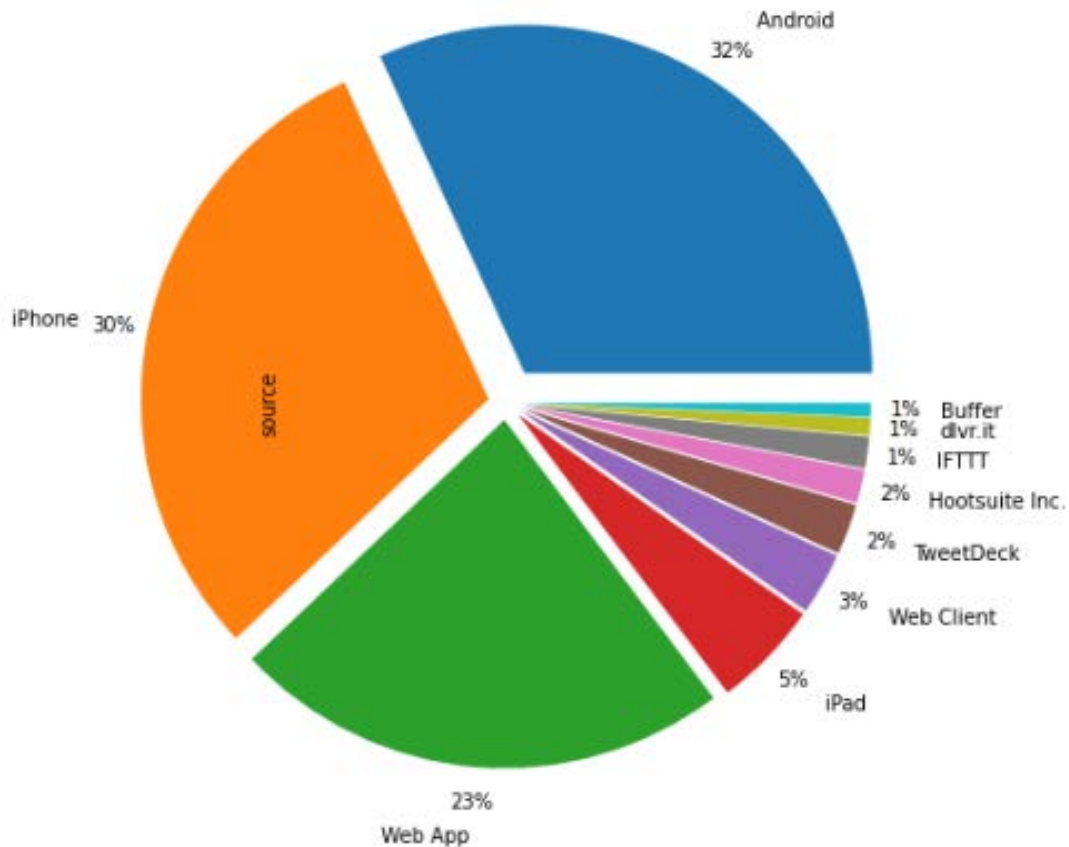


Figure 2- Pie Chart of Tweet Source

I analyzed the top 20 tweet location using word frequencies and they were visualized through word clouds to identify the pandemic hot zones.



Figure 3- Tweet Locations

Figure 3 shows that the pandemic is still an issue worldwide especially in England, United Kingdom and United States. The word cloud also shows frequent tweets about the pandemic from India, Nigeria, this is due to recent spike in cases in these countries.

In addition, I used the Non-Negative Matrix Factorization (NMF) algorithm from the Python sklearn package for topic modelling technique. NMF algorithm has a reputation for being better for learning compact topics. I analyzed popular trends in the United Kingdom (UK) on twitter.

Topics Emerged from Tweets

I grouped identified topics together into 5 board topics.

Topic 1: COVID-19 Origin/Source

There are tweets with keyword like “china”, “investigation origin”, “evidence”, “Wuhan lab”, “originated”.

All these keywords point to discussions about the true origin of this virus and how China might have developed the virus as a bioweapon in a Wuhan laboratory. These keywords also show discussion about China as it is the country where this virus originated from.

Topic 2: COVID-19 Economic Impact on People and Countries

I found tweets with keywords like “household”, “fallen”, “million citizens”, “government”, “files bankruptcy”

These tweets show the negative impacts this pandemic is having on people financially and health wise. A huge number of people filed for bankruptcy, countries economic falling.

Topic 3: COVID-19 Vaccine and Treatment

I found tweets with keywords like “immunity global”, “vaccine”, “pandemic test”, “world country”, “crisis”, “health”, “bill gate”.

These tweets are showing measures the countries of the world and highlights bill gates effort to start vaccine testing for coronavirus. This keywords also shows the topics about health crisis faces many different nations of the world during this pandemic and how many of these countries are struggling to test COVID-19 cases.

Topic 4: Lockdown Ease

I found tweets with keywords like “easing” “restriction”, “Boris”, “lockdown decision”, “ease plan”

These tweets contain topic about plans to ease the lockdown in the UK and measures other countries have been taking to ease the lockdown. The tweets show concerns about the UK easing the lockdown restrictions too early and risking a second wave of the pandemic.

Topic 5: Health Impact on People

Tweets with words like this “worldasthma day”, “emergency”, “asthmauk” were found.

Twitter users in these tweets expressed their fear and stress about the coronavirus due to its quick spread and how similar it is to other respiratory disease like asthma.

Sentiment Analysis of tweets related health impact of coronavirus on people

Sentiment analysis helps us understand people’s feelings towards a specific subject. It is done by identifying, categorizing opinions and transforming the words into actionable insights.

I analyzed tweets for positive, neutral, or negative sentiment. Tweets with a negative sentiment were more common than neutral and positive tweets. The tweets express people concern about the pandemic and how it is affecting their health. Figure 4 shows a plot of the sentiment analysis.

Limitation

I only analyzed English tweets; this could have made me miss out on important trends in countries that do not speak English officially.

There is also limitation with the number of tweets that can be fetched from the twitter API per second, the API also does not allow users to fetch tweets older than a week.

News analysis of COVID-19 on the Economy

Introduction

This project aims to explore news articles related to the economic impact of the COVID-19 pandemic on countries of the world; classification of news articles by sentiment and summarize one of the news articles using a “genism” a python package for summarizing text.

Data Source

The news data source used for this report was from newapi (<https://newsapi.org/>) developer api version. The news are from sources like BBC, TechCrunch, CNN. Figure 6 shows a snippet of the news article.

	Title	Description	Source
0	Technology and ethics in the coronavirus economy	The last two decades have ushered in significant change and transformation. I believe the 2020s will be dispositive in redefining the pillars of our economy, and COVID-19 magnifies this greatly.	TechCrunch
1	Coronavirus: Can live-streaming save China's economy?	In China, the live-streaming industry has become an important platform for economic recovery.	BBC News
2	Sales startup People.ai lays off 18% of staff, raises debt round amid COVID-19 uncertainty	Another startup has turned to downsizing and fund raising to help weather the uncertainty around the economy amid the global coronavirus health pandemic. People.ai, a predictive sales startup backed by Andreessen Horowitz, Iconic, Lightspeed and other investo...	TechCrunch
3	Night burials amid Tanzania's coronavirus defiance	Some fear the East African nation is downplaying the pandemic as it seeks to protect its economy.	BBC News

Figure 6-New Articles

News Scrapping

I queried the newapi for news with keywords combinations like “economy” “covid” “coronavirus”. I extracted title, description and source.

Data Cleaning

I cleaned the data by removing symbols though the data I got from the api was to a very good extent already good to use. I had to covert the data from json to a pandas library dataframe object to allow me carry out the analysis I need to perform on the data easily.

Sentiment Analysis

I performed sentiment analysis on the article obtained from the data source. I performed this analysis to gain insight into how the economy of different countries are reacting to this pandemic.

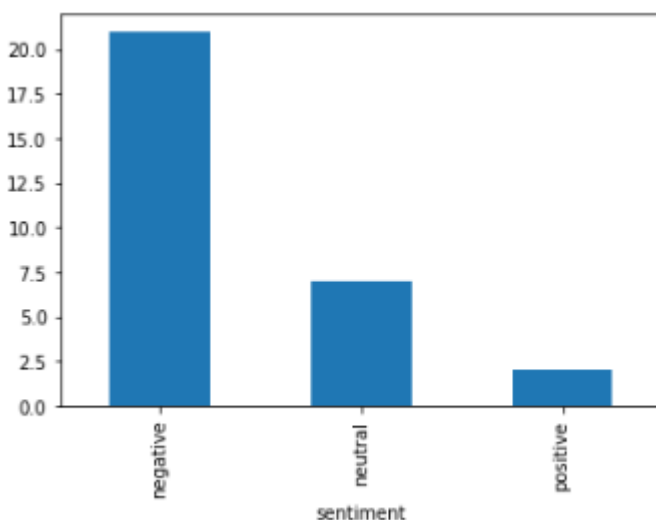


Figure 7 Sentiment Analysis of Economic News

Figure 7 shows an over view of the sentiment in the news article I obtained, as expected they are largely bad news.

I was more interested in seeing the news that are more positive and neural to gain some insights into what makes them different and what country they relate to.

	Title	Description	Source	Sentiment score	sentiment
18	Announcing the Extra Crunch Live event series	The startup world is going through yet another evolution. A few years ago, VCs were focused on growth over profitability. Now, making money is just as important, if not more, than sheer growth. And we're in the midst of a global pandemic, which has brought th...	TechCrunch	0.8658	positive
27	How COVID-19 Contact Tracing Could Help Reopen the Economy - extratv	How COVID-19 Contact Tracing Could Help Reopen the Economy extratv Apple, Google share more details about COVID-19 contact tracing system MobileSyrup How your phone could play a role in coronavirus contact tracing CBC News: The National Citizens Can't Be Forc...	Youtube.com	0.7430	positive

Figure 8 - Positive News

Figure 8 shows in the second news that there might be some hope in times like this. News about contact tracing app been able to help countries open up the economy faster.

I went further to view the neural tweets as well.

	Title	Description	Source	Sentiment score	sentiment
0	Technology and ethics in the coronavirus economy	The last two decades have ushered in significant change and transformation. I believe the 2020s will be dispositive in redefining the pillars of our economy, and COVID-19 magnifies this greatly.	TechCrunch	0.2023	neutral
1	Coronavirus: Can live-streaming save China's economy?	In China, the live-streaming industry has become an important platform for economic recovery.	BBC News	0.2023	neutral
2	Sales startup People.ai lays off 18% of staff, raises debt round amid COVID-19 uncertainty	Another startup has turned to downsizing and fund raising to help weather the uncertainty around the economy amid the global coronavirus health pandemic. People.ai, a predictive sales startup backed by Andreessen Horowitz, Iconic, Lightspeed and other investo...	TechCrunch	0.1027	neutral
11	Hong Kong to Give Free Masks to Every Family As It Prepares to Reopen Economy	Hong Kong's government announced a plan to give every family at least 10 free face masks as it prepares to reopen large parts of society later this week. The region of 7 million people has experienced 14 full days without any community transmission of the new...	Gizmodo.com	0.5106	neutral
14	Answering Your Coronavirus Questions: New Symptoms, Economy And Virtual Celebrations	On this broadcast of The National Conversation, we answer your questions about the economy, mental health and new symptoms of COVID-19. We'll also look at how people are celebrating big life events.	Npr.org	0.5719	neutral
24	Driverless vehicles in the age of the novel coronavirus	The COVID-19 pandemic has led to different outcomes for different businesses. While some have stood to benefit (think Zoom, Facebook and bidet startup Tushy), others have been hit hard and laid off employees in order to survive. But there are some that fall s...	TechCrunch	0.3818	neutral
25	Instant view: Coronavirus ravaged U.S. payrolls in April, 20.5 mln jobs lost	The U.S. economy lost a staggering 20.5 million jobs in April, the steepest plunge in payrolls since the Great Depression and the starkest sign yet of how the novel coronavirus pandemic is battering the world's biggest economy.	Reuters	0.1027	neutral

Figure 9 - Neutral News

It is interesting to see news from China categorized as neutral this means that there is hope for other countries still seriously affected by this pandemic. There is news about tech companies like Facebook and zoom that are benefitting from the pandemic.

Article Summarization

I got selected one of the articles from those I earlier obtained. I used a python library called “BeautifulSoup” to scrap the content of the article.

I used another summarizer library from “genism” to summarize the article in 300 words.

The output of the summary was just having clear as the original article and I was able to get a good over view of what the article was talking about. I experimented by reducing the words to 200 and it was still quite comprehensible for me.

Here is the link to the actual full article:

<https://techcrunch.com/2020/04/17/sales-startup-people-ai-lays-off-18-of-staff-raises-debt-round-amid-covid-19-uncertainty/?guccounter=1>

Here is the summary text:

People.ai, a predictive sales startup backed by Andreessen Horowitz, Iconiq, Lightspeed and other investors and last year valued at around \$500 million, has laid off around 30 people, working out to about 18% of staff, TechCrunch has learned and confirmed.

Alongside that, the company has quietly raised a debt round in the “tens of millions of dollars” to make strategic investments in new products and potentially other moves.

People.ai’s software tracks the full scope of communication touch points between sales teams and customers, supposedly negating the tedious manual process of activity logging for SDRs. The company’s machine learning tech is also meant to generate the average best way to close a deal – educating customer success teams about where salespeople may be deviating from a proven strategy.

People.ai is one of a number of well-funded tech startups that is making hard choices on business strategy, costs and staffing in the current climate.

Some businesses are getting directly hit by the pandemic – for example, house sales and transportation have all but halted, leaving companies in those categories scrambling to figure out how to get through the coming weeks and months and prepare for a potentially long haul of life and consumer and business behavior not looking like it did before January.

But other businesses, like People.ai, which provides predictive sales tools to help salespeople do their jobs better, is (for now at least) falling into that category of IT still in demand, perhaps even more than ever in a shrinking economy.

The company has also just closed a round of debt funding in the “tens of millions” of dollars to use for strategic investments.

Limitations

There is some misclassification in the sentiment analysis.

Conclusion

The COVID-19 pandemic has been affecting many health care systems and nations, claiming the lives of many people and affecting the economy negatively. Twitter projected this heavy toll through the interactions and posts people made related to COVID-19. It is clear that coordinating public health crisis response activities in the real world and online is paramount, and should be a top priority for all health care systems.

We need to build more national and international detection, surveillance and contact tracing systems to detect the spread of infectious diseases.

References

1. Alaa Abd-Alrazaq., et. al “Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study”.
2. Richard J. Medford., et al “An “Infodemic”: Leveraging High-Volume Twitter Data to Understand Public Sentiment for the COVID-19 Outbreak”.
3. <https://www.kaggle.com/nikhilsable/sentiment-using-airline-tweets-using-vader>