**Muma College of Business**

**Independent Study (Spring 2022)**

Eric Massicott, Faith Oluwayemi, Himadeep Movva, Kumar Amrit Vandakara

**Research Topic: Advancing Healthcare through Social Media**

**Objective** – Using social media data to identify preemptive initiatives to improve healthcare/health literacy for the public. The initial topics explored include mental health literacy, substance abuse and COVID-19 misinformation. After consultation with our team’s mentor and some exploration of collectible data for each potential topic, the team made the decision to carry out the study with a focus on tracking COVID-19 misinformation. The objective is to retrieve covid related tweets, perform analysis and address the fake information

**Methodology** – Using APIs and bots, tracking and analyzing health misinformation on social media. After extensive research of data-gathering infrastructure on multiple social media platforms, we planned on extracting data from Twitter.

Data Mining Methodology & Parameters:

|  |
| --- |
| Identify top relevant topics/keywords   * Tweet clusters: cures and treatments, virus characteristics, vaccines, conspiracy theories, politics |
| Binary attributes indicating relevance/irrelevance of tweets |
| Other tweet attributes:   * Tweet ID – a unique identifier for the tweet * User ID – a unique identifier for the user who created the tweet * Time/Location – time tweet was created, location where tweet was created * Text – actual text of the tweet * Hashtags – names of hashtags * Retweets – number of times the tweet has been retweeted * Likes – number of times the tweet has been liked   User attributes   * Follower count – number of followers the user has * Friend count – number of accounts the user follows * Age of account – time the twitter account was created * Tweet count – number of tweets made by the user |

**Milestones met**

* Discovering the research possibilities with data collected from social media
* Examining relevant literature and studying the methodology in similar projects
* Developing Python code and proving a functional API code
* Extracting initial set of data (containing 1000 records)

**Roadblocks crossed**

* We thought to retrieve older data from twitter, but twitter has limitation to access the data prior to seven days. Alternatively, we used to get getoldtweets3 package but couldn’t achieve the desired result
* We also tried to collect the tweets and hydrate the data using twitter hydrator, but we ran into problems pertaining to accessing and hydrating the data
* We aspired to filter, based on the like count, the data but the twitter doesn’t support such feature

**Literature reviewed**

To access the different pieces of literature that we have used thus far, we scanned through the different databases on the USF Libraries Databases. The databases that we specifically used were **Scopus, JSTOR, Google Scholar, and Academic Search Premier.**

When possible, we searched for “covid misinformation” and filtered by the subject area “Computer Science.” We also ordered the literature by most cited and reviewed literature that had been cited at least a total of 30 times.

Our findings led to us finding literature that included information and studies on:

1. Social media posts on covid and distinguishing if the information is factual or false.
2. Collecting data from fact-checking websites.
3. Text Mining tweets from Twitter.
4. Semantic analysis.
5. Different Python packages and how they can be used for our current research.

Our literature are as follows:

* Albahli, S., Algsham, A., Aeraj, S., Alsaeed, M., Alrashed, M., Tayyab Rauf, H., Arif, M., & Abed Mohammed, M. (2021). Covid-19 public sentiment insights: A text mining approach to the Gulf Countries. *Computers, Materials & Continua*, *67*(2), 1613–1627.
* Shahi, G. K., Dirkson, A., & Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on Twitter. *Online Social Networks and Media*, *22*, 100104.
* Gupta R., Vishwanath A., Yang Y. (2021). COVID-19 twitter datasets with latent topics, sentiments, and emotions attributes.
* Buntain, C., & Golbeck, J. (2017). Automatically identifying fake news in popular Twitter threads. *2017 IEEE International Conference on Smart Cloud (SmartCloud)*.
* Sharma K., Seo S., Meng C., Rambhatla S., Liu Y. (2020). COVID\_19 on Social Media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.
* Gruzd, A., & Mai, P. (2020). Going viral: How a single tweet spawned a COVID-19 conspiracy theory on Twitter. *Big Data & Society*, *7*(2), 2053951720938405.