

Guidelines for Case Project

Social Media Analytics - CS-EJ5621, Fall 2020

The main objective of the case project is to obtain hands-on experience of carrying out analysis on real-life social media dataset. In the case project, you will go through different stages of social media analytics including data gathering, cleaning, analysis, interpretation, and presenting the outcomes. This project will also support you in envisioning and carrying out future projects based on social media data.

You are required to submit a brief proposal by **02.10.2020**. The case project proposal should briefly describe the background and motivations of carrying out the analysis on the chosen Twitter hashtag(s) or use account(s). Furthermore, what kind of analysis you plan to carry out, and how it will support in answering your research question(s).

It is highly recommended to use TAGS for data collection, but if you are familiar/have used any other tool (or would like to build your coding capabilities in Python/R) and would like to use it instead, consult the course instructor. Most of the analysis can be carried out in Google sheets (Microsoft Excel), TAGS, and other GUI tools introduced during the lecture. Alternatively, you are free to choose any other coding or online/offline tool(s) deemed fit for your needs.

For the final projects, I would recommend to focus on healthcare related content on Twitter (*though the students can freely choose other domain/topics/platform of their choice*). Following are some of the examples of highly active healthcare related hashtags and user accounts. In order to have useful insights and findings, I would strongly recommend choosing an active topic or user account (adequate data stream of few hundred tweets per day) on Twitter that interests you personally.

Active Twitter hashtags	Active Twitter accounts
#ADHD	@HEAS_Vic
#Diabetes	@CDCemergency
#PublicHealth	@WebMD
#Smoking	@CDC_eHealth
#DigitalHealth	@FDATobacco
#Alcohol	@Act4HlthyKids
#VaccinesWork	@muscle_fitness
#Asthma	@ActiveSchoolsUS

The analysis and final deliverable can potentially cover (but not limited to) the following:

1) Descriptive analysis

- Total number of tweets and retweets
- Temporal and geographical analysis
- Users analysis (e.g. most active users, influencers)
- Hashtags analysis (unique hashtags, frequently recurring hashtags)
- URLs analysis (unique URLs and frequently recurring URLs)

2) Content Analysis

- Tweets categorization of 300 randomly selected tweets (e.g. informational, promotion, resource, etc).
- Profiles analysis of users getting highest number of retweets (top 20).
- Sentiment analysis of the analytical sample.

3) Social Network analysis

For SNS, use the TAGS explorer feature in TAGS. You can also use Gephi or other SNA tool.

The sequence of stages together with the deadlines in the project process are:

Project Proposal (due 02.10.2020)

In the proposal (maximum 2 pages), briefly describe the dataset (selected hashtags or user accounts) you are intending to collect and analyze. Briefly, describe your motivations behind the selected hashtag or accounts. Who can benefit from this analysis, and why? What are the tools you will use for collection, analysis, visualization etc.? Any impending factors that you foresee.

Project deliverable (due 20.11.2020)

Submit your case project final deliverable (**e.g. report, presentation, video etc.**) by the deadline. Although you are free to choose the reporting format, however, it should essentially cover the following topics:

- Brief background about the selected hashtag(s) or user accounts(s)
- Significance of studying the topic
- Data collection method, timeframe, and tools
- Data analysis method and tools
- Results and findings
- Implications of the findings
- Challenges/lessons learned

References

- [Characterizing the followers and tweets of a marijuana-focused Twitter handle](#)
- [Tweeting for and Against Public Health Policy: Response to the Chicago Department of Public Health's Electronic Cigarette Twitter Campaign](#)
- [Characterizing vaping posts on Instagram by using unsupervised machine learning](#)
- [Diabetes on Twitter: A Sentiment Analysis](#)
- [Activity, content, contributors, and influencers of the twitter discussion on urologic oncology](#)
- [The social dynamics of lung cancer talk on Twitter, Facebook and Macmillan.org.uk](#)
- [On the use of multimedia in Twitter health communication: analysis of tweets regarding the Zika virus](#)
- [Diffusion of Messages from an Electronic Cigarette Brand to Potential Users through Twitter](#)