Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220562

Automated Twitter Extraction and Visual Analytics with Dashboards: Development and First Experimentations

Haris PATEL[a], Riyâz PATEL[a], Djamel ZITOUNI[b], Benjamin GUINHOUYA[a,b], Mathilde FRUCHART[a,b] and Antoine LAMER[a,b,1]
[a]*Univ. Lille, Faculté Ingénierie et Management de la Santé, F-59000, Lille, France*
[b]*Univ. Lille, CHU Lille, ULR 2694 - METRICS: Évaluation des Technologies de santé et des Pratiques médicales, F-59000 Lille, France*

Abstract. Information found in the social media may help to set up infoveillance and track epidemics, identify high-risk behaviours, or assess trends or feelings about a subject or event. We developed a dashboard to enable novice users to easily and autonomously extract and analyze data from Twitter. Eleven users tested the dashboard and considered the tool to be highly usable and useful. They were able to conduct the research they wanted and appreciated being able to use this tool without having to program.

Keywords. Data Science; Social media, Visual analytics; Education

1. Introduction

Information found in social media may help to set up infoveillance and track epidemics, identify high-risk behaviours, or assess trends or feelings about a subject or event [1,2]. In this study, we developed a dashboard to enable novice users to easily and autonomously extract and analyze data from Twitter. We assessed their ability to use it and captured their feedback on the value of such a tool for their daily practice.

2. Methods

The extraction of Twitter data is carried out using the open-source python package *snscrape*. The application is developed with *Dash*, an open-source framework for building web interactive applications entirely in Python. The dashboard allows the user to create a dataset, without coding, according to four criteria: the text query, the desired time range, the number of tweets to be retrieved and the language of the tweets.

The dashboard provides three tabs with (i) navigation across the dataset, and its exportation in CSV format, (ii) a time analysis with the count of tweets for each day and for each hour of the day, and (iii) a text analysis with a word

[1] Corresponding Author, Antoine Lamer, ULR 2694, 2 place de Verdun, F-59000, Lille, France; E-mail: antoine.lamer@univ-lille.fr

frequency graph and a word cloud [3]. We have proposed this tool to people who do not master programming (physicians, pharmacists and community managers), by instructing them to conduct a research with a keyword related to their fields of interest.

3. Results

Eleven users tested the dashboard (for topics such as supervised injection site, launching of a French prevention phone number about suicide, health policies, etc.) and considered the tool to be highly usable and useful. The figure 1 displays the wordcloud of frequency terms for query about supervised injection site. It highlights the towns in which supervised injection site are open or in the process of opening.

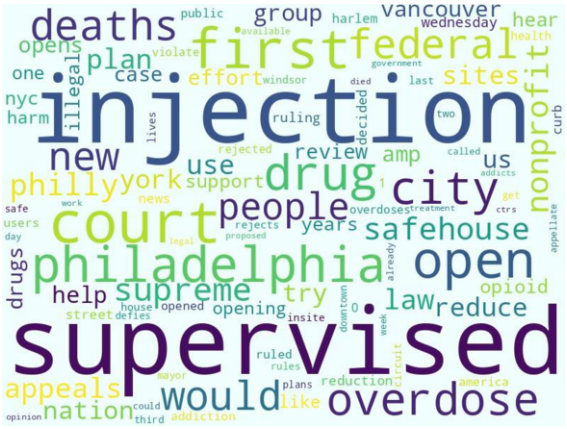

Figure 1. Wordcloud in the text analysis panel, for tweets about supervised injection site

4. Discussion and Conclusions

We offer a free tool which automatically extracts tweets and provides descriptive statistics. It can be used to follow the trends of current topics discussed on Twitter. With its search module, there is no need to program. The tool is available online and can be used for free.

References

[1] Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study, J. Med. Internet Res. 22 (2020). doi:10.2196/19016.

[2] Castillo-Sánchez G, Marques G, Dorronzoro E, Rivera-Romero O, Franco-Martín M, De la Torre-Díez I. Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review, J. Med. Syst. 44 (2020) 205. doi:10.1007/s10916-020-01669-5.

[3] Twitter Analytics Dashboard, (n.d.). https://dash-ilis22.herokuapp.com/ (accessed January 26, 2022).