Internship: Understanding users profiles by analyzing massive non structured conversations on a drug users web community

Advisors: Benoit Favre (LIS, AMU), Raquel Urena (Sesstim, AMU)

Context

Today, digital technologies constitute promising solutions to meet the challenges facing the French healthcare system (Moquet-Anger, 2020). In the field of addictions, systems for managing and analyzing massive amounts of data are already playing a major role in redefining clinical and public health research practices (Thierry et Reynaud, 2019). Nowadays, advances in AI such as social network analysis and natural language processing offer novel opportunities for better understanding patterns of drug use and their associated consequences. By analyzing the massive data existing thanks to naturally occurring, self-disclosed communication on social media, it is possible to understand temporal trends of problematic drug use, market changes, social norms and cultural aspects of drug use and relevant psychological factors (Kim et al., 2017).

Objectives

In spite of storing information from almost a million of users, psychoatif.org lacks structured metadata describing users demographics, habits, goals and problems. The objective of this internship is to infer user profiles from the textual information published and read by the users in their posts, as well as their interactions in the virtual community. This internship takes place in the context of a larger project, AI4DU, which will exploit the generated profiles in order to better understand the community and eventually help the community better face drug-related problems.

Scientific program

This research proposal will be using the website psychoactif.org that provides drug users with an anonymous platform to exchange knowledge on drug use, practices, harm reduction strategies and prevention. Psychoactif.org is the largest French speaking web community of drug users with approximately 1 million users, 3 million page views per month and between 200 and 400 new testimonial messages created per day.

AI based analyses including, natural language processing and topic modeling to obtain a structured user profile with information about drug use, behaviors, interests and prevention measures will be carried out. The intern will work with large language models to detect fine-grained pieces of information, thanks to annotations gathered by public health researcher specialists in addictions. She/he will also build user representations from their behavior, comparing deep methods with classical topic modeling methods. The generated profiles will be validated with the help of SanteRCom and psychoactif.org researchers through semi-structured interviews with the user base.

This is a multidisciplinary internship in which the intern will have the opportunity to deal with a real research project involving public health researchers, psychologists and data scientists.

Additional information

Skills: Master-level computer science, an interest for linguistics, python programming, deep learning, Pytorch, rigor and tenacity, a hunch for research, excellent communication skills.
Location: the internship will take place between the Sesstim on the Faculty of Medicine campus, and the LIS at Luminy campus of Aix-Marseille University.
Dates: Spring-summer 2023, duration 5-6 months.
Wages: regulatory internship salary (about 500 euros/month).
Computation: the intern will have access to the Jean-Zay GPU cluster for running experiments.
Contact details : To apply please send a mail with your CV, a motivation letter and note sheets to raquel.urena@univ-amu.fr and benoit.favre@lis-lab.fr .