Utilizing Topic Modelling for Assessing Mhealth App's Risks to Users' Health before and during the COVID-19 Pandemic

Authors : Pedro Augusto Da Silva E Souza Miranda, Niloofar Jalali, Shweta Mistry

Abstract : BACKGROUND: Software developers utilize automated solutions to scrape users' reviews to extract meaningful knowledge to identify problems (e.g., bugs, compatibility issues) and possible enhancements (e.g., users' requests) to their solutions. However, most of these solutions do not consider the health risk aspects to users. Recent works have shed light on the importance of including health risk considerations in the development cycle of mHealth apps to prevent harm to its users. PROBLEM: The COVID-19 Pandemic in Canada (and World) is currently forcing physical distancing upon the general population. This new lifestyle made the usage of mHealth applications more essential than ever, with a projected market forecast of 332 billion dollars by 2025. However, this new insurgency in mHealth usage comes with possible risks to users' health due to mHealth apps problems (e.g., wrong insulin dosage indication due to a UI error). OBJECTIVE: These works aim to raise awareness amongst mHealth developers of the importance of considering risks to users' health within their development lifecycle. Moreover, this work also aims to help mHealth developers with a Proof-of-Concept (POC) solution to understand, process, and identify possible health risks to users of mHealth apps based on users' reviews. METHODS: We conducted a mixed-method study design. We developed a crawler to mine the negative reviews from two samples of mHealth apps (my fitness, medisafe) from the Google Play store users. For each mHealth app, we performed the following steps: • The reviews are divided into two groups, before starting the COVID-19 (reviews' submission date before 15 Feb 2019) and during the COVID-19 (reviews' submission date starts from 16 Feb 2019 till Dec 2020). For each period, the Latent Dirichlet Allocation (LDA) topic model was used to identify the different clusters of reviews based on similar topics of review The topics before and during COVID-19 are compared, and the significant difference in frequency and severity of similar topics are identified. RESULTS: We successfully scraped, filtered, processed, and identified health-related topics in both qualitative and quantitative approaches. The results demonstrated the similarity between topics before and during the COVID-19.

Keywords : natural language processing (NLP), topic modeling, mHealth, COVID-19, software engineering, telemedicine, health risks

Conference Title : ICDHR 2021 : International Conference on Digital Healthcare and Research
Conference Location : Berlin, Germany
Conference Dates : May 20-21, 2021