Forensic science international: Digital investigation

Digital forensics supported by machine learning for the detection of online sexual predatory chats

C.H. Ngejane[ab], J.H.P. Eloff[a], T.J. Sefara[b], V.N. Marivate[c]

[a] Cyber-security & Big Data Science Research Group Department of Computer Science, University of Pretoria, South Africa

[b] Council for Scientific and Industrial Research, Pretoria, South Africa

[c] Data Science for Social Impact Research Group Department of Computer Science, University of Pretoria, South Africa

https://www.sciencedirect.com/science/article/pii/S2666281721000032

Abstract

Chat-logs are informative digital footprints available on Social Media Platforms (SMPs). With the rise of cybercrimes targeting children, chat-logs can be used to discover and flag harmful behaviour for the attention of law enforcement units. This can make an important contribution to the safety of minors on SMPs from being exploited by online predators. The problem is that digital forensic investigation is mostly manual. Thus, a daunting task for forensic investigators because of the sheer volume and variety of data. The solution that is proposed in this paper employs a Digital Forensic Process Model that is supported by Machine Learning (ML) methods to facilitate the automatic discovery of harmful conversations in chat-logs. ML has already been successfully applied in the domain of text analysis for the discovery of online sexual predatory chats. However, there is an absence of approaches that show how ML can contribute to a digital forensic investigation. Thus, the contribution of this paper is to indicate how the tasks in a digital forensic investigation process can be organised so to obtain useable ML results when investigating online predators.