Citizen-generated data and machine learning: an innovative methode to study violence against women

Claudia Villante, Gianpiero Bianchi, Alessandra Capobianchi and Maria Giuseppina Muratore (Italian Institute of National Statistics (Istat), Italy)

claudia.villante@istat.it

Abstract

From 2018 to date ISTAT collects and releases data of 1522 helpline made by Department for Equal Opportunities of the Presidency of the Council of Ministers to support and help victims of gender-based violence and stalking. Collecting and using this citizen-generated data on violence against women is of the outmost importance because complements data collected through administrative sources and dedicated surveys, by providing timely and accurate information, gathered by NGO's skilled professionals of the service, according also with the technical UN Women/WHO guidance "Improving the collection and use of administrative data on violence against women". The objective of the paper is to provide guidance on how to collect, process, including with machine learning techniques data from 1522, focusing on the adopted technique of using textual data, from the transcription of calls. The paper presents the motivations, methodological choices and techniques adopted to use an essential and non- traditional data source for understanding the phenomenon of violence against women and to contribute within the multi- sources approach initiated by ISTAT for an increasingly accurate and in-depth study of the phenomenon. The impact that the adopted machine learning methodologies have had on improving quality in the data collection process is also described.