Let-Mi: An Arabic Levantine Twitter Dataset for Misogynistic Language

Hala Mulki
The ORSAM Center
for Middle Eastern Studies
Ankara, Turkey
hala.mulki@orsam.org.tr

Bilal Ghanem
University of Alberta
Edmonton, Canada

bilalhgm@gmail.com

Abstract

Misogyny is one type of hate speech that disparages a person or a group having the female gender identity; it is typically defined as hatred of or contempt for women. Online misogyny has become an increasing worry for Arab women who experience gender-based online abuse on a daily basis. Such online abuse can be expressed through several misogynistic behaviors which reinforce and justify underestimation of women, male superiority, sexual abuse, mistreatment, and violence against women. Misogyny automatic detection systems can assist in the prohibition of anti-women Arabic toxic content. Developing these systems is hindered by the lack of the Arabic misogyny benchmark datasets. In this work, we introduce an Arabic Levantine Twitter dataset for Misogynistic language (LeT-Mi) to be the first benchmark dataset for Arabic misogyny[1] The proposed dataset consists of 6,550 tweets annotated either as neutral (misogynistic-free) or as one of seven misogyny categories: discredit, dominance, cursing/damning, sexual harassment, stereotyping and objectification, derailing, and the threat of violence. We further provide a detailed review of the dataset creation and annotation phases. The consistency of the annotations for the proposed dataset was emphasized through interrater agreement evaluation measures. Moreover, Let-Mi was used as an evaluation dataset through binary,multi-class, and target classification tasks which were conducted by several state-of-the-art machine learning systems along with Multi-Task Learning (MTL) configuration. The obtained results indicated that the performances achieved by the used systems are consistent with state-of-the-art results for languages other than Arabic, while employing MTL improved the performance of the misogyny/target classification tasks.

---

[1] https://github.com/bilalghanem/let-mi