1. Data

The data set available in this paper is composed of 2285 definitions gathered from Urban Dictionary [2]. After a classification process done by three independent researchers, 1034 definitions were classified as misogynistic and 1251 as non-misogynistic.

2. Experimental design, materials, and methods

All data were downloaded using Urban Dictionary API [3]. The original data set was composed of 2,606,521 definitions posted by 2,001,482 distinct users between the launch of the Urban Dictionary
platform in 1999 and May 2016. After a filtering process (described next), the data set available in this paper was created and it is composed of 2285 definitions.

A list of 51 words typically associated with misogynistic content was created by one researcher with extensive domain knowledge. This bag of words was used to filter 951,978 potentially misogynistic definitions out from the initial data set. The rationale behind pre-filtering the data set is that potentially misogynistic words can be used in non-misogynistic sentences too (e.g. “ass fucking lesbians” as misogynistic; “misspelling of lesbian” as non-misogynistic).

A sample of 2285 definitions was extracted and manually classified by two independent researchers. Disagreements were resolved by a third one. Out of those 2285 definitions, 1034 were classified as misogynistic and 1251 as non-misogynistic.

2.1. Ethical considerations

The data set was anonymized before its publication. Warning: This data set contains material that many will find offensive.

Acknowledgments

This work was partly funded by the World Technology Universities Network (WTUN) and the Irish Institute of Digital Business (dotLAB).

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.104223.

References

[1] T. Lynn, P.T. Endo, P. Rosati, I. Silva, G.S. Leoni, D. Ging, A Comparison of Machine Learning Approaches for Detecting Misogynistic Speech in Urban Dictionary, Social Media, 2019.
[2] Urban dictionary. Available at: https://www.urbandictionary.com/. Last access: March, 2019.
[3] Urban dictionary API. Available at: https://github.com/mattbierner/urban-dictionary-entry-collector. Last access: March, 2019.