



ELSEVIER

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Benchmark dataset of memes with text transcriptions for automatic detection of multi-modal misogynistic content

Francesca **Gasparini**, Giulia **Rizzi**, Aurora **Saibene**, Elisabetta **Fersini***

Department of Informatics, Systems and Communication, University of Milano-Bicocca, Italy

ARTICLE INFO

Keywords:

Misogyny detection
Multi-modal content
Memes
Cybersexism
Visual and textual cues

ABSTRACT

In this paper we present a benchmark dataset generated as part of a project for automatic identification of misogyny within online content, which focuses in particular on memes. The benchmark here described is composed of 800 memes collected from the most popular social media platforms, such as Facebook, Twitter, Instagram and Reddit, and consulting websites dedicated to collection and creation of memes. To gather misogynistic memes, specific keywords that refer to misogynistic content have been considered as search criterion, considering different manifestations of hatred against women, such as body shaming, stereotyping, objectification and violence. In parallel, memes with no misogynist content have been manually downloaded from the same web sources. Among all the collected memes, three domain experts have selected a dataset of 800 memes equally balanced between misogynistic and non-misogynistic ones. This dataset has been validated through a crowdsourcing platform, involving 60 subjects for the labelling process, in order to collect three evaluations for each instance. Two further binary labels have been collected from both the experts and the crowdsourcing platform, for memes evaluated as misogynistic, concerning aggressiveness and irony. Finally for each meme, the text has been manually transcribed. The dataset provided is thus composed of the 800 memes, the labels given by the experts and those obtained by the crowdsourcing validation, and the transcribed

<http://dx.doi.org/10.1016/j.dib.xxxx.xx.xxx>

2352-3409/© 2021 The Author(s). Published by Elsevier Inc.

texts. This data can be used to approach the problem of automatic detection of misogynistic content on the Web relying on both textual and visual cues, facing phenomenons that are growing every day such as cybersexism and technology-facilitated violence.

Specification Table

Subject	Artificial Intelligence
Specific subject area	Automatic detection of multi-modal content
Type of data	Tables (.csv) Images (.jpeg)
How data were acquired	Images of memes were downloaded from the Web and saved in high quality JPG format with maximum dimension resized to 640 pixels in order to have a dataset of normalized dimensions; labelled data were collected from domain experts and through a crowdsourcing platform; text data were manually transcribed.
Data format	Raw data Analyzed data (labeled and transcribed)
Parameters for data collection	Memes were downloaded from popular social media platforms and websites dedicated to the collection and creation of memes, using keywords related to misogyny. The memes were subsequently manually selected. Each meme was evaluated by three domain experts and by three subjects through a crowdsourcing platform.
Description of data collection	The 800 memes of the dataset were selected from a bigger collection, by a pool of three experts of the domain in order to be balanced between misogynistic and non-misogynistic ones. The dataset has been validated by using the Figure Eight crowdsourcing platform (now called Appen, https://appen.com/), involving 60 subjects, and the text of each meme has been manually transcribed.
Data source location	Raw data as well as crowdsourcing labels were collected from the Web Data were processed at: Institution: University of Milano-Bicocca City: Milan Country: Italy
Data accessibility	In a public repository: Repository name: github Data identification name: Misogynistic-MEME Direct URL to data: https://github.com/MIND-Lab/MEME Instructions for accessing these data: Data are password protected. Password will be provided after fulfilling a copyright notice.
Related research article	E. Fersini, F. Gasparini, S. Corchs, Detecting sexist MEME on the Web: A study on textual and visual cues. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). IEEE, 2019. p. 226-231.

*Corresponding author: Tel.: +39-02-6448-7896.

e-mail: francesca.gasparini@unimib.it (Francesca Gasparini), g.rizzi10@campus.unimib.it (Giulia Rizzi), a.saibene2@campus.unimib.it (Aurora Saibene), elisabetta.fersini@unimib.it (Elisabetta Fersini)

Value of the Data

- The role of women within the society has increased in importance and the way we approach and refer to them is crucial. Misogyny is a form of discrimination towards women and has been spreading exponentially through the Web. A popular communication tool in social media platforms are memes which are typically composed of pictorial and textual components, and that are often used to convey misogynistic messages. Automatic detection of misogynistic content is becoming fundamental to counteract online discrimination, cybersexism and violence. As misogyny is conveyed by both textual and visual media, a dataset of multimodal content, such as a dataset of memes, is mandatory to build efficient machine learning techniques that promptly intercept these offensive messages online.
- The presented dataset will be particularly useful for: i) researchers from Natural Language Processing and Computer Vision communities that deal with social media data analysis and that are interested in developing machine learning models that combine textual and visual cues; ii) companies that develop artificial intelligent strategies to control social media activities; and iii) social science researchers that study gender discrimination especially on the Web.
- This dataset presents a data structure that can be adopted for the collection of further data on the topic to develop more robust machine learning techniques. In fact the automatic detection of misogynistic content is particularly challenging especially considering that: (i) misogynistic and non misogynistic memes can share the same visual content but a different text, and (ii) misogyny can be expressed by text, image or by their combination. This dataset can also be adopted to increase the quality of the labelling task, as it can be used as a gold standard to check the reliability of annotators on crowdsourcing platforms.
- This dataset will pave the way not only to the development of machine learning techniques able to promptly intercept on offensive content against women the Web, but also to stimulate other researchers in devoting their attention to this phenomenon, increasing the awareness and sensitivity to misogyny as well as to other forms of discrimination.
- This dataset provides the labels from both domain experts and subjects recruited within the population. The comparison between the two distributions is also significant, not only with respect to misogyny by itself, but also with respect to the perception of aggressiveness and irony.
- Up to our knowledge, this is the first dataset of memes that face the misogynist phenomenon. A related dataset is the one proposed by Facebook AI for the Hateful Meme (HM) Challenge [1]. The main difference between the HM dataset and the proposed one is that our memes have been collected from social media platforms, describing a real scenario, while the others could refer to synthetic memes automatically generated as benign confounders of similar hateful memes. Moreover, while the HM dataset is generally devoted to general hateful and non-hateful memes, without focusing on any specific types and targets of hate, our dataset is centered on women as target.

Data Description

In the latest years, misogyny has found in the Web a new and powerful way of diffusion together with the new phenomenon of cybersexism, where women are often victim of offensive messages and, in the most serious cases, of abuse and threats. Online platform providers have

introduced policies to prevent offensive content. However, due to the speed of dissemination of messages in social media, systems able to automatically filter offensive content are urgently needed [2, 3]. While new opportunities for females have been opened on the Web, systematic inequality and discrimination offline is replicated in online spaces in the form of offensive contents against them [4, 5].

Memes are especially popular communication means for social media users, being able to efficaciously convey funny and/or ironic jokes [6]. A meme is defined as an image composed of a pictorial information on which a text is superimposed a posteriori by a human [6]. Given the characterization of memes, they have been progressively used to convey hate [7] and, in this specific dataset, we are interested in memes circulating in the Web and targeting women with sexist and aggressive messages [8, 9].

In order to develop efficient machine learning techniques able to automatically detect multi-modal misogynistic messages online, we here present a dataset composed of:

- **800 memes**, saved as jpeg images, resized to have the greatest dimension equal to 640 pixels. These memes are saved with a progressive unique ID.
- A **table** saved as a .csv file, where all the data collected are reported, according to the following structure:
 - *memeID*: unique identifier associated to the meme;
 - *text*: transcription of the text reported in the meme;
 - *misogynisticDE*: Boolean attribute related to the presence of misogynistic content as reported by the Domain Experts (DE);
 - *aggressiveDE*: Boolean attribute; in case of a misogynist meme it represents the presence of aggressiveness, as reported by the DE;
 - *ironicDE*: Boolean attribute; in case of a misogynist meme it represents the presence of irony, as reported by the DE;
 - *misogynisticCS*: Boolean attribute related to the presence of misogynistic content, as reported by the annotators of the CrowdSourcing platform (CS);
 - *aggressiveCS*: Boolean attribute; in case of misogynist meme it represents the presence of aggressiveness, as reported by the CS;
 - *ironicCS*: Boolean attribute; in case of misogynist meme it represents the presence of irony, as reported by the CS.
 - *confidence_M_CS*: agreement on the misogynist attribute among the CS;
 - *confidence_A_CS*: agreement on the aggressiveness attribute among the CS;
 - *confidence_I_CS*: agreement on the misogynist irony among the CS;

We underline that the labels of the 3 experts have an agreement of 100% for all the 800 memes, as it was the criterion to select them among all the downloaded memes.

Data Distribution. The distribution of the three labels (misogyny, aggressiveness and irony) given by the domain experts are reported in Figure 1. As the dataset of 800 memes was selected starting from the misogynistic DE labels, the first pie chart on the left confirms that the dataset is equally distributed within the two classes with 400 memes each. The second and third pie charts

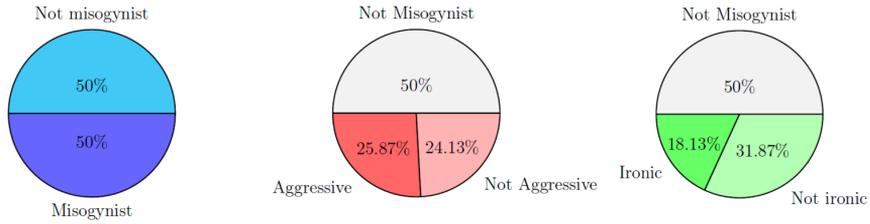


Fig. 1. Distribution of the three labels given by the domain experts (DE)

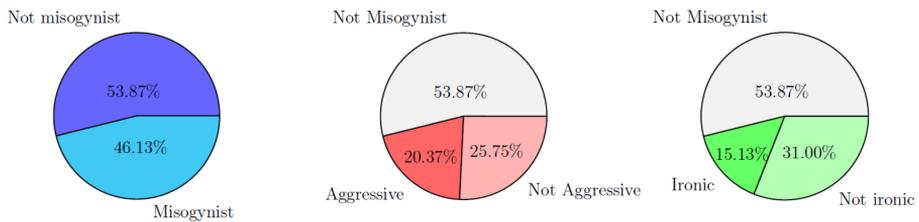


Fig. 2. Distribution of the three labels given by the crowdsourcing annotators (CS)

reported for the memes labelled as misogynistic the percentage of them considered respectively aggressive and ironic.

In Figure 2, the corresponding label distributions given by the CS annotators are reported. The first pie on the left shows that in this case the memes labelled as misogynistic are less than in the case of the DE evaluation.

Given the labels provided by DE and CS, 59 memes have been annotated differently with respect to the misogynistic label. Among them, 76.28% were considered misogynistic by the experts but not by the annotators, while only 23.72% of them were considered misogynistic only by the annotators. The complexity of the annotation process is also reflected by the agreements of the CS labellers, which have been reported in the corresponding columns in the Annotation Data sheet.

Experimental Design, Materials and Methods

Meme collection process

The most popular social media platforms, i.e. Facebook, Twitter, Instagram and Reddit, have been considered in the data collection phase. Memes that convey potential misogynistic content have been collected in the October-November 2018 period through the following operations [10]:

- Searching for threads or conversations dedicated and written by anti-women/feminists supporters, such as the Men Going Their Own Way (MGTOW) website and the related thread on Reddit;
- Exploring discussions on sexism in political or social events;

- Browsing hastags such as #girl, #girlfriend, #women.

Subsequently, the dataset has been enlarged by collecting memes from websites dedicated to meme creation and/or collection, as follows:

- Browsing hashtags such as #girl, #girlfriend, #women, #feminist;
- Consulting collections on all the variations of famous memes involving female characters.

In parallel, memes with non-misogynistic contents have been manually downloaded from the same web sources and by adopting the same keywords, for a non trivial collection of the memes.

Expert labelling and dataset definition

Three domain experts have evaluated all the collected memes, labelling them as misogynist or non-misogynist. The final dataset is composed of 800 memes, selected among those with an agreement of 100%, in order to have a perfect balanced dataset with respect to the two classes. The experts have also annotated the misogynistic memes with respect to aggressiveness and irony. This phase provided the three corresponding boolean labels reported in the data sheet with DE suffix.

Dataset Annotation through the crowdsourcing platform

The 800 memes selected were labelled adopting a crowdsourcing platform (Figure Eight in 2018, now called Appen, <https://appen.com/>).

A controlled labelling experiment was chosen to provide judgments from trusted and reliable participants with equally distributed age (between 20 and 50 years old) and gender. All the data were anonymized.

The annotation task was designed as follows:

- the order of the memes in the experiment is randomized to avoid bias;
- the maximum number of judgments that any contributor can provide is limited to 40 memes, to limit fatigue and consequent unreliable annotations;
- any annotator can leave whenever he/she wants;
- for each annotator, the task expires after an hour and a half, regardless of the number of evaluated memes, in order to limit external stimuli;
- each annotation page shows only one meme at a time, to not influence or bias the participant by seeing other meme contemporaneously;
- each meme is evaluated by three different subjects.

For each meme, the question **In your opinion, is this meme misogynistic?** is proposed to the participant, as depicted in Figure 3.

Then, only in the case of a meme evaluated as misogynistic, the following two questions were proposed:

1. In your opinion, is this meme ironic?
2. In your opinion, is this meme aggressive?



In your opinion, is this meme misogynistic? (required)

- Yes
 No

Fig. 3. The first question in the crowdsourcing annotation task

Definitions and guidelines were not provided to the CS annotators, wanting to collect their perception of misogyny, irony and aggressiveness present in the meme, without influencing their decision. This crowdsourcing annotation provided the three corresponding boolean labels reported in the data sheet with CS suffix, together with the corresponding confidence level (in terms of percentage of agreement).

Text transcription

Finally, the text superimposed to the imaged have been manually transcribed for each of the 800 memes.

Ethics Statement

An informed consent was given to all the involved subjects, indicating the presence of possible explicit contents, explaining how to perform the tasks also reporting them the possibility to stop the labeling activity whenever they wanted. Any personal data have been acquired to identify the labellers, being therefore GDPR compliant.

Acknowledgments

We want to give our thanks especially to Silvia Corchs for her valuable scientific support in facing the problem of misogyny detection, and to Monica Mantovani and Gaia Campisi, for their supporting work during the collection and validation of the dataset.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

References

- [1] D. Kiela, H. Firooz, A. Mohan, V. Goswami, A. Singh, P. Ringshia, D. Testuggine, The hateful memes challenge: Detecting hate speech in multimodal memes, 2021. [arXiv:2005.04790](https://arxiv.org/abs/2005.04790).
- [2] M. Anzovino, E. Fersini, P. Rosso, Automatic identification and classification of misogynistic language on twitter, in: International Conference on Applications of Natural Language to Information Systems, Springer, 2018, pp. 57–64.
- [3] F. Gasparini, I. Erba, E. Fersini, S. Corchs, Multimodal classification of sexist advertisements., in: ICETE (1), 2018, pp. 565–572.
- [4] S. Frenda, B. Ghanem, M. Montes-y Gómez, P. Rosso, Online hate speech against women: Automatic identification of misogyny and sexism on twitter, Journal of Intelligent & Fuzzy Systems 36 (2019) 4743–4752.
- [5] F.-M. Plaza-Del-Arco, M. D. Molina-González, L. A. Ureña-López, M. T. Martín-Valdivia, Detecting misogyny and xenophobia in spanish tweets using language technologies, ACM Transactions on Internet Technology (TOIT) 20 (2020) 1–19.
- [6] L. Shifman, Memes in a Digital World: Reconciling with a Conceptual Troublemaker, Journal of Computer-Mediated Communication 18 (2013) 362–377.
- [7] T. Farrell, M. Fernandez, J. Novotny, H. Alani, Exploring misogyny across the manosphere in reddit, in: Proceedings of the 10th ACM Conference on Web Science, 2019, pp. 87–96.
- [8] M. Paciello, F. D’Errico, G. Saleri, E. Lamponi, Online sexist meme and its effects on moral and emotional processes in social media, Computers in Human Behavior 116 (2021) 106655.
- [9] M. A. Franks, Unwilling avatars: Idealism and discrimination in cyberspace, Colum. J. Gender & L. 20 (2011) 224.
- [10] E. Fersini, F. Gasparini, S. Corchs, Detecting sexist meme on the web: A study on textual and visual cues, 2019, pp. 226–231. doi:[10.1109/ACIIW.2019.8925199](https://doi.org/10.1109/ACIIW.2019.8925199).