

# Identifying Sexist Content in Tweets



## 1 TASK 1: Sexism Identification in Tweets

The first subtask is a binary classification. The systems must decide whether a given tweet contains sexist expressions or behaviours (i.e., it is sexist itself, describes a sexist situation or criticizes a sexist behaviour), and classify it according to two categories: YES and NO.

- Examples of sexist tweets ("YES") are:
  - "It's less of #adaywithoutwomen and more of a day without feminists, which, to be quite honest, sounds lovely."
  - "I'm sorry but women cannot drive, call me sexist or whatever but it is true."
  - "You look like a whore in those pants" - My brother of 13 when he saw me in a leather pant.
- Examples of not sexist tweets ("NO") are:
  - "Where all the white women at?"
  - "The shocking video of a woman at the wheel who miraculously escapes an assassination attempt"

## 2 TASK 2: Source Intention in Tweets

The second subtask is a multi-class classification. For the tweets that have been predicted as sexist, the second task aims to classify each tweet according to the intention of the person who wrote it. One of the three following categories must be assigned to each sexist tweet:

- DIRECT: The intention is to write a message that is sexist itself, as in:
  - "A woman needs love, to fill the fridge, if a man can give this to her in return for her services (housework, cooking, etc.), I don't see what else she needs".

- **REPORTED:** The intention of the author is to report or describe a sexist situation or event suffered by a woman or women in first or third person, as in:
  - "I doze in the subway; I open my eyes feeling something weird: the hand of the man sat next to me on my leg #SquealOnYourPig".
- **JUDGEMENTAL** The intention of the author is to be judgemental since the tweet describes sexist situations or behaviors with the aim to condemning them.
  - "As usual, the woman was the one quitting her job for the family's welfare".

### 3 RULES

You need to submit the model predictions in files named `Team.Task1.TryNb.csv` and `Team.Task2.TryNb.csv`. Each file should contain two columns: ID and PRED. The PRED column represents the model's predictions as an estimated distribution of the annotators' votes. For example, in TASK 1, `[1/6 5/6]` indicates 1 YES vote and 5 NO votes; in TASK 2, the prediction `[2/6 3/6 1/6]` means 2 votes DIRECT, 3 votes REPORTED, and 1 vote JUDGEMENTAL.

You will not have access to the test set votes, which are used to evaluate your model. The evaluation criterion is the Macro F1 score, which is the arithmetic mean of the F-scores for each class. To compute the F-score on soft predictions, we use the method outlined in Erhani, Johan, et al. "Confusion Matrices: A Unified Theory." IEEE Access (2024).

Rules are:

- You can submit up to 3 files per task to optimize your ranking;
- Model predictions must sum to 1 to be evaluated; otherwise, they will be replaced by a zero vector;
- To ensure fairness, the use of LLMs (large language models) or transformers that require GPUs is not allowed.

Good luck to all!