**SOURCE**

Train dataset was collected from Kaggle [Jigsaw/Conversation AI](https://www.kaggle.com/jigsaw-team) challenge **(**[**https://www.kaggle.com/c/jigsaw-unintended-bias-in-toxicity-classification/data**](https://www.kaggle.com/c/jigsaw-unintended-bias-in-toxicity-classification/data)**)** as a Zip file.

**SIZE**

CSV zip file size 273 mb (Uncompressed file size is 778 MB)

File contains around **1.8 million entries**

**Feature of dataset**

* It’s a labelled dataset with 45 columns.
* It contains columns like comment, **target**, severe\_toxicity, obscene, insult, threat, male, female etc.
* Model will be able to learn from the **target** column and predict the toxicity level of comment between 0 and 1, where 0 being **Not toxic** and 1 being **very toxic.**
* Dataset also includes various identities mentioned in the comments like Male, Female, Transgender, Other gender, No gender etc.