**Dataset Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Type** | **Reference** | **Additional Notes:** |
| English Twitter Data Set | Tweets annotated for hate speech | Waseem and Hovy, NAACL 2016; Waseem, NLP and CSS 2016  https://github.com/zeerakw/hatespeech |  |
| German Twitter Data Set | Tweets regarding refugees in Germany; annotated with hate speech ratings | Ross et al. NLP4CMC 2016  https://github.com/UCSM-DUE/IWG\_hatespeech\_public | could be used as foreign language comparison? |
| Wikipedia Abusive Language Data Set | Discussion comments from English Wikipedia; annotated on whether it contains personal attack + demographic data for crowd who annotated | Wikipedia Detox project  Wulczyn, Thain N., and Dixon L. (2017). “Ex machina: Personal attacks seen at scale”, in *ICWWW*, pp. 1391–1399.  https://figshare.com/articles/Wikipedia\_Detox\_Data/4054689  <https://arxiv.org/abs/1610.08914>  <https://meta.wikimedia.org/wiki/Research:Detox/Data_Release>  <https://github.com/ewulczyn/wiki-detox/blob/master/src/figshare/Wikipedia%20Talk%20Data%20-%20Getting%20Started.ipynb> | Over 100k comments |
| WikiConv | Mulitlingual corpus encompassing history of conversations on Wikipedia Talk | Hua, Danescu-Niculescu-Mizil C., Taraborelli D., Thain N., Sorensen J., and Dixon L., (2018). “Wikiconv: A corpus of the complete conversational history of a large online collaborative community”, *arXiv preprint.*  https://arxiv.org/pdf/1810.13181.pdf  https://convokit.cornell.edu/documentation/wikiconv.html  https://github.com/conversationai/wikidetox/tree/master/wikiconv | includes deletion, modification, restoration of comments |
|  | 2 datasets based on Wikipedia Talk pages; include context | Pavlopoulos J., Sorensen J., Dixon L., Thain N., and Androutsopoulos I. (2020). “Toxicity Detection: Does Context Really Matter?”, in *Proc. of 58th Annual Meeting of Association for Computational Linguistics,* pp. 4296-4305. | 1st has 250 comments |