Information Retrieval Approaches

This paper makes a comparative study on various methods to detect cyberbullying and sarcasm on social media. The problem in the previous work in cyberbullying detection is the low of accuracy and the number of false positive also they cannot detect sarcasm along with cyberbullying.

Dataset

The previous work in cyberbullying detection used different datasets. [Walisa Romsaiyud ,etal](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Walisa%22&searchWithin=%22Last%20Name%22:%22Romsaiyud%22&newsearch=true)[1] used from two different datasets the first one is the posted messages by members in Perverted-justice used as training datasets, and the second one is Twitter datasets from Stanford University as testing datasets. [Sani Muhamad Isa](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Sani%20Muhamad%22&searchWithin=%22Last%20Name%22:%22Isa%22&newsearch=true), etal [2] and Vikas S Chavan etal [3] used a textual conversation taken from the Kaggle ([www.kaggle.com](http://www.kaggle.com)) which provides 1,600 conversations in Formspring.me. Harsh Dani,etal [5] used dataset from twitter and Myspace labeled as normal or bullying. Maral Dadvar,etal[4] collect his dataset from haresment comments other misbehavioursof YouTube videos [Paras Dharwal](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Paras%22&searchWithin=%22Last%20Name%22:%22Dharwal%22&newsearch=true)etal[6] ,Edwin Lunando etal[7] , MONDHER BOUAZIZI etal[8] and [S.K.Bharti](https://www.sciencedirect.com/science/article/pii/S235286481630027X" \l "!)etal [9] in their papers of sarcasm detection they collected the dataset manually from Twitter.

Differences Among the datasets

The main difference in the datasets is the datasets from social m.edia from social media donnot have ground truth but the date sets from websites like kaggle .com have ground truth .