**Project Updates**

|  |  |  |
| --- | --- | --- |
| **Year** | **Research Paper** | **Dataset** |
| 2020 | [Aggressive, Repetitive, Intentional, Visible, and Imbalanced: Refining Representations for Cyberbullying Classification](https://arxiv.org/abs/2004.01820) | [Dataset - Cyberbullying](https://github.com/cjziems/cyberbullying-representations/tree/master/data) |
| 2020 | [Can We Achieve More with Less? Exploring Data Augmentation for Toxic Comment Classification](https://arxiv.org/pdf/2007.00875.pdf) | [Dataset - Wikipedia](https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/discussion)\* |
| 2019 | [A Machine Learning Approach to Comment Toxicity Classification](https://arxiv.org/abs/1903.06765) | [Dataset - Wikipedia](https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/discussion)\* |
| 2018 | [Convolutional Neural Networks for Toxic Comment Classification](https://dl.acm.org/doi/10.1145/3200947.3208069) | [Dataset - Wikipedia](https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/discussion)\* |
| 2018 | [Challenges for Toxic Comment Classification: An In-Depth Error Analysis](https://arxiv.org/pdf/1809.07572.pdf) | [Davidson hate speech dataset](https://github.com/t-davidson/hate-speech-and-offensive-language)  [Dataset - Wikipedia](https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/discussion)\* |
| 2017 | [Automated Hate Speech Detection and the Problem of Offensive Language](https://arxiv.org/pdf/1703.04009.pdf) | [Davidson hate speech dataset](https://github.com/t-davidson/hate-speech-and-offensive-language) |

* We are going to mainly use the highlighted dataset.
* Description: Wikipedia talk pages presented by Google Jigsaw during Kaggle’s Toxic Comment Classification Challenge

Imp links

<https://github.com/deepandas11/HAN-and-Data-Augmentation-Text-Classifier>

<https://medium.com/analytics-vidhya/hierarchical-attention-networks-d220318cf87e>

<https://github.com/synthesio/hierarchical-attention-networks/blob/master/model.py>

<https://www.kaggle.com/sanghan/attention-with-fasttext-embeddings>