Detecting Toxic Comments using CNN

**Abstract:**

The increase in penetration of usage of internet services has increased exponentially in the past 4 months due to the ongoing pandemic, this has enabled a large number of active new and old users using the internet for various services ranging from academic, entertainment, industrial, monitoring and the emergence of a new trend in the corporate-life i.e work-from-home. Due to this sudden emergence of the crowd using the internet, there has been a rise in the number of mischievous persons too. Now it is the primary task of every online platform provider to keep the conversations constructive and inclusive. The best example can be referred to, can be twitter, a social media platform where people share their views. This platform has already drawn a lot of flak because of the spread of hate speech, insults, threat, defamatory acts which becomes a challenge for many such online providers in regulating them. Thus, there is active research being conducted in the field of Toxic comment classification. Here we compare different deep learning and other trivial techniques on the dataset and propose a model that outperforms all others and compares them one-on-one. We have undertaken the Kaggle dataset for the above purpose which has been widely used and one of the prime resources for scholars working in deducing the challenge of toxic comment classification.The results would help up to develop an online interface where we would be able to identify the toxicity level in the given phrase or sentence and classify them into their order of toxicity.