**Multi Class Classification Using BERT**

For this task I have used data from the Kaggle competition toxicity classification. I have finetuned DISTILBERT on this data.

Data have 6 types of categories namely – toxic, severe toxic, obscene, threat, insult, identity hate. Number of records is 159, 571. It has 0 null values.

First I split the data and identify the target column, then I tokenised it using then DISTILBERT tokenizer. Also the model that I am using is DISTILBERT-BASE-UNCASED. After this I cleaned the data using the CLEAN module from CLEANTEXT. After cleaning the data, I encoded the sentences using the tokenizer.encode\_plus class, it convert the data into input id and the masks which is stored in a list and then pass it to the model for the training purposes.

The model has a embedding layer, a global max pool 2d layer, two dense layer and dropout.

The model is compiled using adam optimizer, categorical cross entropy as the loss function and accuracy as the performance matrix.

The model is trained on a single epoch and a batch size of 64.