

# Toxic Tweets Dataset: NLP Problem

This dataset has a collection of Tweets. It's labelled as Toxic - 1, Non-toxic - 0. Apply the NLP methods to predict the toxicity of the tweets. Download the dataset from the following

## Kaggle Compitation

<https://www.kaggle.com/datasets/ashwiniyer176/toxic-tweets-dataset>.

All the credits to the original collectors.

## Procedure:

1. Convert the CSV file to the panda data frame.
2. Convert the text to the following.
  - Bag of Words
  - TF-IDF
3. For the obtained features, apply the following methods of prediction.
  - Decision Trees
  - Random forest
  - Naive Bayes Model
  - K-NN Classifier
  - SVM
4. For each of this method produce the following metrics:
  - Precision, Recall, F1 - Score
  - Confusion Matrix
  - RoC - AUC curve