Problem Statement: Hate Speech Classification

Dataset using Twitter data, is was used to research hate-speech detection. The text is classified as: hate-speech, offensive language, and neither. Due to the nature of the study, it’s important to note that this dataset contains text that can be considered racist, sexist, homophobic, or generally offensive.

**Column Description:**

**count**

number of CrowdFlower users who coded each tweet (min is 3, sometimes more users coded a tweet when judgments were determined to be unreliable by CF)

**hate\_speech**

number of CF users who judged the tweet to be hate speech

**offensive\_language**

number of CF users who judged the tweet to be offensive

**neither**

number of CF users who judged the tweet to be neither offensive nor non-offensive

**class**

class label for majority of CF users. 0 - hate speech 1 - offensive language 2 - neither

**tweet**

text tweet

**Target Column:**

Need to predict the class column for test data by applying suitable NLP based algorithms.

**class**

class label for majority of CF users. 0 - hate speech 1 - offensive language 2 - neither