Group Name - Arjohi
Name - Hiten Chadha
Email - hitenchadha1995@gmail.com
Country - Denmark
College/Company - Technical University of Denmark
Specialization - NLP

Problem Description:

The term hate speech is understood as any type of verbal, written or behavioural communication that attacks or uses derogatory or discriminatory language against a person or group based on what they are, in other words, based on their religion, ethnicity, nationality, race, colour, ancestry, sex or another identity factor. In this problem, We will take you through a hate speech detection model with Machine Learning and Python.

Hate Speech Detection is generally a task of sentiment classification. So for training, a model that can classify hate speech from a certain piece of text can be achieved by training it on data that is generally used to classify sentiments. So for the task of hate speech detection model, We will use the Twitter tweets to identify tweets containing Hate speech.

Data Understanding:

We will analyze a dataset CSV file from Kaggle containing 31,935 tweets. The dataset was heavily skewed with 93% of tweets or 29,695 tweets containing nonhate labeled Twitter data and 7% or 2,240 tweets containing hate-labeled Twitter data. We will try different classification algorithms after the preprocessing and data cleaning steps.

Number of NA values: 0

Outliers: NA

Skewed: Skewed/Imbalanced class

Approach to deal with NA values, outliers, skewed data etc.:

In this case, since our dataset is an array of tweets which are essentially strings, we do not face any outliers or any skewed data as such. To deal with any missing data, if present, we deal it with the SimpleImputer class in sklearn.impute library.

For dealing with the unbalanced class, we can use the resampling technique. It consists of removing samples from the majority class (under-sampling) and/or adding more copies from the minority class (over-sampling). Since, we don’t have a large dataset, oversampling can be a better choice for this task.