Details

Name	Email	Country	College/Company	Specialization
Kelvin Mpofu	mpofukelvintafadzwa@gmail.com	South Africa	n/a	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 a 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.

Github Repo Link

https://github.com/mpofukelvintafadzwa/Data_Glacier_NLP_INTERNSHIP/tree/main/Week_109

EDA performed on the data

1. Qualitative Analysis

    • Average word length vs label

    • label counts

    •count of words distribution

2. Quantitative Analysis

    • Mutual Information scores for categorical features, measures mutual dependence between the categorical features and the target variable.

    • Up sampling the minority class

    • Created word clouds for text analysis.

    •Compared unsampled labels to the sampled labels

    .

Final Recommendation

Only those features which are found to be significant should be used when building machine learning models.